

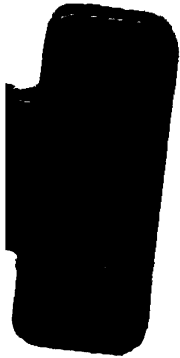
UC-NRLF



B 3 958 636



CHEMISTRY



PUNCHED CARDS

THEIR APPLICATIONS TO SCIENCE AND INDUSTRY

SECOND EDITION

Edited By

ROBERT S. CASEY

*W. A. Sheaffer Pen Co.
Fort Madison, Iowa*

JAMES W. PERRY

*Center for Documentation and Communication Research
Western Reserve University
Cleveland, Ohio*

MADELINE M. BERRY

*National Science Foundation
Washington, D. C.*

And

ALLEN KENT

*Center for Documentation and Communication Research
Western Reserve University
Cleveland, Ohio*

REINHOLD PUBLISHING CORPORATION

NEW YORK

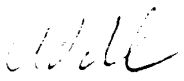
CHAPMAN & HALL, LTD., LONDON

Copyright 1958 by
REINHOLD PUBLISHING CORPORATION

All rights reserved

Library of Congress Catalog Card Number: 58-12824

REINHOLD PUBLISHING CORPORATION
Publishers of Chemical Engineering Catalog, Chemical
Materials Catalog, "Automatic Control," "Materials in
Design Engineering," "Progressive Architecture"; Ad-
vertising Management of the American Chemical Society


CHEMISTRY

PRINTED IN THE U.S.A.

2180
A1C33
1958
CHEMISTRY
LIBRARY

PREFACE TO THE SECOND EDITION

The editors are pleased with the reception given the first edition of this book. Our readers were apparently more tolerant of its shortcomings than were the editors.

For this edition the purposes remain the same as stated in the preface to the first edition, namely: to help the individual worker, to record present knowledge and experience, and to discuss some general principles as a stimulus to further development.

Some idea of the rapidity with which the field has grown may be gained from the fact that the bibliography of uses contains 400 entries, compared with 276 entries in the first edition. This great increase is reflected in the extension of the Practical Applications Section (Part II) from 186 pages in the first edition to 295 pages in the present book. Here the reader will find a broad survey of such important and unique uses as the Peek-a-Boo System; the Uniterm System; mechanized coding and searching techniques applied to the metallurgical literature; the Zato-coding System; and a most interesting discussion of the use of punched cards in linguistic analysis as applied to ancient texts such as the Dead Sea Scrolls.

The general plan of presentation for Parts I to V is the same as in the first edition. The following chapters are unchanged: 4. Preparing Reports, Papers and Books; 20. Correlation of Research Data; 21. Mathematical Analysis of Coding; 26. Searching the Literature; 17. Plant Breeding and Genetics has been lengthened slightly. The other chapters were rewritten or replaced.

The chapter on Computation was eliminated. The potentialities of punched-card machines for scientific computation are so vast that they cannot be covered, or even adequately outlined within the limitations of a chapter. The bibliography, Chapter 30, contains references to applications of punched-card computations.

Substantial advances in the science and art of punched-card applications have been recorded since the first edition. Subject matter analysis has been receiving the attention of numerous workers, and much progress has been reported. This is perhaps the most important topic in the punched-card field and like most other topics in that field, is still in need of further development. In certain limited areas advancements in the punched-card art are becoming more definitive; for example: Qualitative Analysis by Spectral Methods; Metallurgical Literature; Classification, Searching and Mechanization in the U. S. Patent Office; and Library Applications.

Chapters 7 and 15 are representative case histories of installations of commercial systems which have had successful applications.

The applications of punched cards are too extensive to report all outstanding work. It is possible in a single book to pick only cases which are representative of different fields and different methods. To illustrate the diversity of applications, Chapter 14 was included, which gives miniature case histories in a variety of fields.

For beginners, unfamiliar with punched cards, the following chapters are recommended as an introductory survey: Chapter 1; in Chapter 2, p. 12, pp. 18–23 (Elementary Subject Matter Analysis and Coding), and pp. 27–29 (Supplementary Notes on Hand-Sorted Cards); Chapter 4 illustrates the use of a simple bibliographic punched-card file. Skim through Chapter 14, Review of Applications.

The editors are grateful for all the help received in preparing this second edition. They are particularly grateful to the authors of the chapters, and to friends who called their attention to punched-card references. R. S. C. thanks his employer, the W. A. Sheaffer Pen Company, for the use of the company's facilities, and his secretary, Mrs. Dorothy Billman, for typing and other help.

October 1, 1958

ROBERT S. CASEY
JAMES W. PERRY
MADELINE M. BERRY
ALLEN KENT

PREFACE TO THE FIRST EDITION

In the present phase of our scientific age, a situation has developed in which research "publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships."¹ The very bulk of the rapidly expanding mass of scientific and technical information threatens to impair the usefulness of scientific investigation.

The tendency of accumulations of scientific and technical information to become unwieldy is evident even in files of very modest size; even with small files of information, dissatisfaction has developed with the results obtained from conventional tools, such as ordinary file cards, classified report files, etc. It has been discovered that considerable improvement in speed and ease of locating information in files of modest size can be achieved by using punched cards of simple type, viz., edge-punched cards sold in the United States under the trademarks "Keysort," "E-Z Sort" and "Rocket" and in England as "Paramount" or "Cope-Chat" cards.

This book is directed principally to the needs of the individual scientist, engineer, or other technologist, whether in the laboratory, field, industrial plant, library, school or executive office. Our primary purpose is to furnish sufficient information to permit the application of punched-card techniques to individual problems. However, the present state of knowledge of this subject does not allow full definitive treatment. There are many scattered bits of information about punched-card applications, each indicating the value of mechanical aids to the solution of intellectual problems. But, this knowledge needs to be extended and correlated. Many of the procedures described are preliminary, tentative and experimental.

Therefore, another purpose of this book is to record present knowledge and experience so that better use of the presently available punched-card devices, and design of devices better suited to practical needs, will be stimulated. In addition, some general principles are discussed which may also apply to types of mechanical devices not yet invented.

The hand-sorted edge-punched cards are discussed in greater detail than the machine-sorted cards. In fact, one object has been to make the book serve as an operating instruction manual for the edge-punched cards. It is not possible to do the same for machine-sorted cards within the scope of this book. The editors feel that familiarity with the easily learned sorting

¹ Bush, Vannevar, *Atlantic Monthly*, 176, 101-8 (July, 1945).

procedures and techniques of using hand-sorted cards will facilitate understanding of the more elaborate sorting procedures possible with machine-type cards. The machine-sorted cards and the machines for manipulating them are described and illustrated. The emphasis is on their *use*. Various applications and case histories are reported. The full possibilities of their applications are so vast that it is not possible to set down in one place all operating instructions. Detailed directions for each application must be obtained through consultation with technicians from the manufacturers.

The editors do not imply preference for one type of card over another: on the contrary, they hope that this book will help to evaluate the shortcomings and the advantages of different types of punched cards for different purposes.

Part I is introductory and elementary. It contains sufficient information to permit an individual to set up and use a simple punched-card file.

Part II consists of case histories of punched-card applications carefully selected to show what has already been accomplished. The most effective utilization of punched-card methods is learned more readily from practice and experience than from rule and precept. For this reason, Part II forms the heart of this book.

Part III is more general and theoretical in nature. Fundamental problems involved in applying punched-card techniques to intellectual activities, and vice versa, are discussed. Some of the chapters of Part III consider the general problem of organizing information quite apart from punched-card techniques. It is hoped that these chapters may contribute generally to the advance of the art of information analysis and also stimulate further investigation of the possibilities inherent in punched-card techniques.

Part IV is a study, speculative in nature, as to the role that punched cards and related devices may eventually play in relationship to other methods for coping with information problems.

It is not possible at present to define the realm of usefulness of punched-card techniques. It is, of course, obvious that punched-card techniques will supplement rather than supplant existing methods in handling information. No revolutionary schemes on a large scale are advocated at present.

Part V is a bibliography on uses of punched cards in connection with scientific information. Papers on the subject are so widely scattered that a résumé of this sort appeared advisable. The editors will appreciate having their attention directed to any pertinent papers that may have been overlooked.

As the chapters have been written by different authors, a certain amount of overlapping and repetition is inevitable. The editors feel that this is not wholly undesirable in view of the present undeveloped state of the art. Furthermore, the editors have made no effort to reconcile differences in

opinion among the various authors. It is not always possible to be certain what is irreconcilable difference in opinions and what is merely difference in viewpoint. The editors feel that this treatment is to the advantage of the discriminating reader.

The editors solicit suggestions for improvement of future editions.

Editing this book has been a pleasant task, thanks, first of all, to the generous cooperation of the authors of the various individual chapters. Thanks are also due the American Chemical Society, whose Board of Directors, through its Committee on Punched Cards, has supported and encouraged the study of punched-card techniques for chemical information problems. We are grateful to the punched-card companies whose products are mentioned throughout the book, for the use of illustrations and for much helpful technical information. Miss Madeline M. Berry contributed skillful assistance in checking the manuscript and reading proof. We also wish to thank Miss Alice M. Perry, who typed most of the manuscript.

Completion of our editorial task within a reasonable time would scarcely have been possible without financial support accorded one of us (J. W. P.) by the Carnegie Foundation through the Center for Scientific Aids to Learning at M. I. T. The other editor (R. S. C.) thanks his employer, the W. A. Sheaffer Pen Co., for use of the company's facilities, and his secretary, Mrs. Dorothy Billman, for much typing and other assistance.

JAMES W. PERRY
Cambridge, Mass.
ROBERT S. CASEY
Fort Madison, Iowa

June, 1951

CONTENTS

Chapter	Page
PREFACE TO SECOND EDITION.....	iii
PREFACE TO FIRST EDITION.....	v

PART I: FUNDAMENTAL MACHINE CONSIDERATIONS

1. INTRODUCTION, <i>Allen Kent, James W. Perry, and Robert S. Casey</i>	3
2. ELEMENTARY MANIPULATIONS OF HAND-SORTED PUNCHED CARDS, <i>Robert S. Casey and James W. Perry</i>	12
3. COMMERCIALY AVAILABLE EQUIPMENT AND SUPPLIES, <i>Thomas H. Rees, Jr.</i> ..	30

PART II: PRACTICAL APPLICATIONS OF PUNCHED CARDS AND RELATED DEVICES

4. THE USE OF PUNCHED-CARD TECHNIQUES IN PREPARING REPORTS, PAPERS, AND BOOKS, <i>Charles A. Burkhard</i>	93
5. AN INTERNATIONAL CLASSIFICATION AND PUNCHED-CARD FILING SYSTEM FOR METALLURGICAL LITERATURE, <i>Marjorie R. Hyslop and Alvina Wassenberg</i> . 100	
6. THE PEEK-A-BOO SYSTEM—OPTICAL COINCIDENCE SUBJECT CARDS IN INFORMATION SEARCHING, <i>W. A. Wildhack and Joshua Stern</i>	125
7. A UNITERM SYSTEM FOR REPORTS, <i>Charles E. Zerwekh, Jr.</i>	152
8. AN IMPROVED ANESTHESIA RECORD CARD, <i>Max S. Sadove and Myron J. Levin</i> . 161	
9. PUNCHED CARDS AS AIDS TO QUALITATIVE CHEMICAL ANALYSIS BY SPECTRAL METHODS, <i>L. E. Kuentzel</i>	175
10. AN APPLICATION OF RANDOM CODES FOR LITERATURE SEARCHING, <i>Claire K. Schultz</i>	232
11. SEARCHING METALLURGICAL LITERATURE, <i>Allen Kent and James W. Perry</i> ..	248
12. CLASSIFICATION, SEARCHING AND MECHANIZATION IN THE U. S. PATENT OFFICE, <i>B. E. Lanham and J. Leibowitz</i>	261
13. APPLICATION OF PUNCHED CARDS TO LIBRARY ROUTINES, <i>Madeline M. Berry</i> . 279	
14. REVIEW OF APPLICATIONS, <i>Barbara L. Haksteen</i>	303
15. A CASE HISTORY OF A ZATOCODING INFORMATION RETRIEVAL SYSTEM, <i>Claude W. Brenner and Calvin N. Mooers</i>	340
16. THE USE OF PUNCHED CARDS IN LINGUISTIC ANALYSIS, <i>Rev. Roberto Busa</i> ... 357	
17. AN ABSTRACTING AND INFORMATION SERVICE FOR PLANT BREEDING AND GENETICS, <i>R. H. Richens</i>	374

PART III: FUNDAMENTAL CONSIDERATIONS IN CODING AND SYSTEMS DESIGN

18. SUBJECT MATTER ANALYSIS AND CODING—SOME FUNDAMENTAL CONSIDERATIONS, <i>James W. Perry</i>	391
19. HOLES, PUNCHES, NOTCHES, SLOTS, AND LOGIC, <i>C. D. Gull</i>	422
20. CORRELATION OF RESEARCH DATA AND ESTABLISHMENT OF CAUSE AND EFFECT RELATIONSHIPS, <i>D. E. H. Freear</i>	432
21. MATHEMATICAL ANALYSIS OF CODING SYSTEMS, <i>Carl S. Wise</i>	438
22. COMPREHENSIVE CODING SCHEMES FOR CHEMICAL COMPOUNDS, <i>D. E. H. Freear</i>	465
23. SUPERIMPOSED CODING WITH THE AID OF RANDOMIZING SQUARES FOR USE IN MECHANICAL INFORMATION SEARCHING SYSTEMS, <i>H. P. Luhn</i>	492
24. INDEXING AND INDEX SEARCHING, <i>E. J. Crane and Charles L. Bernier</i>	510
25. MAKING CLASSIFICATION SYSTEMS FOR PUNCHED-CARD CODING, <i>Norman T. Ball</i>	528
26. SEARCHING THE LITERATURE, <i>Byron A. Soule</i>	542
27. TRANSCRIPTION PROBLEMS IN PREPARING AND USING PUNCHED-CARD FILES, <i>C. D. Gull</i>	555

PART IV: FUTURE POSSIBILITIES

28. EVALUATION OF MECHANIZED DOCUMENTATION AT THE GMELIN INSTITUTE, <i>E. H. Eric Pietsch</i>	571
29. DATA PROCESSING MACHINES FOR FILING AND RETRIEVING TECHNICAL INFORMATION, <i>Ascher Opler</i>	619

PART V: ANNOTATED BIBLIOGRAPHY ON USES OF PUNCHED CARDS

30. ANNOTATED BIBLIOGRAPHY ON USES OF PUNCHED CARDS, <i>Robert S. Casey</i>	638
AUTHOR INDEX	673
SUBJECT INDEX	679

Part I

FUNDAMENTAL MACHINE CONSIDERATIONS

Chapter 1

INTRODUCTION

ALLEN KENT AND JAMES W. PERRY

Western Reserve University, Cleveland, Ohio

AND

ROBERT S. CASEY

W. A. Sheaffer Pen Co., Fort Madison, Iowa

Methods and systems for expediting the recall and correlation of recorded information by applying various mechanical and electronic devices have made rapid strides during the past decade. The "recorded information" may be entries in notebooks, records on pieces of paper, correspondence files, collections of reprints, notes on file cards, accounts, financial transactions, measurements, calculations, pictures, diagrams, drawings, descriptions of people and things—almost anything the human mind can conceive. "Recorded information" on one hand represents all the books and journals ever printed; on the other hand, it is the growing accumulation of data in your own laboratory or office. It is toward solution of the latter problem that this book is directed.

Punched cards are being applied to a steadily widening range of subject matter. One result has been to stimulate interest in similar applications of various electronic devices, especially computers, and another has been to initiate the development of specially designed searching and selecting machines. At present, however, a variety of punched cards is the most widely used type of mechanical aid for facilitating the retrieval and correlation of recorded information.

The two general types of punched cards, for hand sorting and machine sorting, have been in the process of development for almost two centuries. The control card for looms, invented in 1780 by Joseph Jacquard, laid the foundation for the future development of information storage tools. The loom control card stored the information necessary to reproduce patterns consistently during the weaving of fabrics.

Another pioneering development of unusual importance in the development of present-day information control devices was the "analytical engine" invented by Charles Babbage about 1840. This device used pre-punched cards to facilitate statistical control.

The first punched cards and equipment for manipulating them which resembled modern counterparts appeared about 1880, when Dr. Herman

Hollerith introduced a pantograph punch and an electric accounting tabulator with sorting box for use in connection with the United States Census. The commercial organization formed by Hollerith, The Tabulating Machine Company, which merged in 1911 into the Computing-Tabulating-Recording Company, was the forerunner of International Business Machines Corporation and paved the way for other companies entering this field (see chapter 3).

Hand-sorted punched cards as an aid to "preventing the accidental misplacement of a card in the files" made an elementary appearance in 1904 with a system which required the notching of the bottom edges of cards according to the file section in which the cards were to be placed. Rods in a suitable holder at the bottom of the card tray forced incorrectly filed cards (notches in wrong place) to pop up when placed in any but the correct section of the file.¹ Another invention in 1907, based upon a variation of this principle, had lifting-bars at the bottom of the card tray which could be inserted in appropriate positions in order to select the desired cards which were notched at various positions along the bottom edge of the card.² These developments appear to have been the forerunners of the hand-sorted punched cards commercially available today.

The What, Why and How of using punched cards are not as obvious as they are in the case of wrenches, hammers, screw drivers and other simple tools. Some idea of the What and Why of punched cards is given in this chapter.

As mentioned earlier, the two general types of punched cards in common use are hand-sorted and machine-sorted. The hand-sorted type has one, two or more rows of holes along one or more edges of the card. Meanings are assigned to individual holes or to combinations of holes. The holes on a given card, appropriate to the entries to be punched on that card, are clipped open to the edge of the card, forming notches as shown in Figure 1-1. The sorting needle or "tumbler" resembles a single-tine ice pick with a blunt point or a knitting needle with a handle. When it is inserted in a given hole in a group of cards and lifted, the cards on which that hole has been notched drop from the pack (Figure 1-2).

Two different types of machine-sorted cards are illustrated in Figure 1-3. In the IBM card, the twelve punching positions in a vertical column constitute a coding unit. Digits are indicated by punching one of the positions number 0-9, while for each of the letters two holes are punched in the same column. In Remington-Rand cards, the twelve holes in each column are divided into two sets of six holes each. Each set of holes is a coding unit. In the punching system used by Remington-Rand, meaning is attached to

¹ U. S. Patent No. 759,483, to W. K. Sparrow (May 10, 1904).

² U. S. Patent No. 873,305, to E. Eckart (Dec. 10, 1907).



Figure 1-1. Punching a notch in the edge of a hand-sorted card.

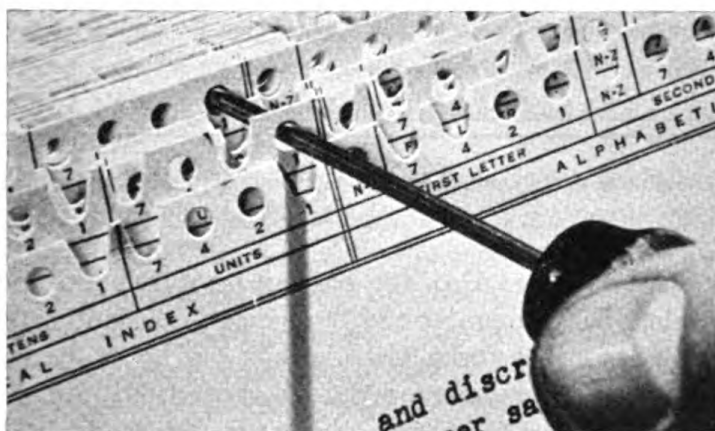


Figure 1-2. Hand sorting a file of edge-punched cards.

the punching of a single hole in a set and to the punching of combinations of two and three holes. For details see Figure 1-3.

The cards are mechanically punched and in actual use are fed automatically through machines in which the punched holes cause electrical or mechanical contacts to be made, thus actuating mechanisms which perform the desired operations. If desired, all cards coded in a given manner

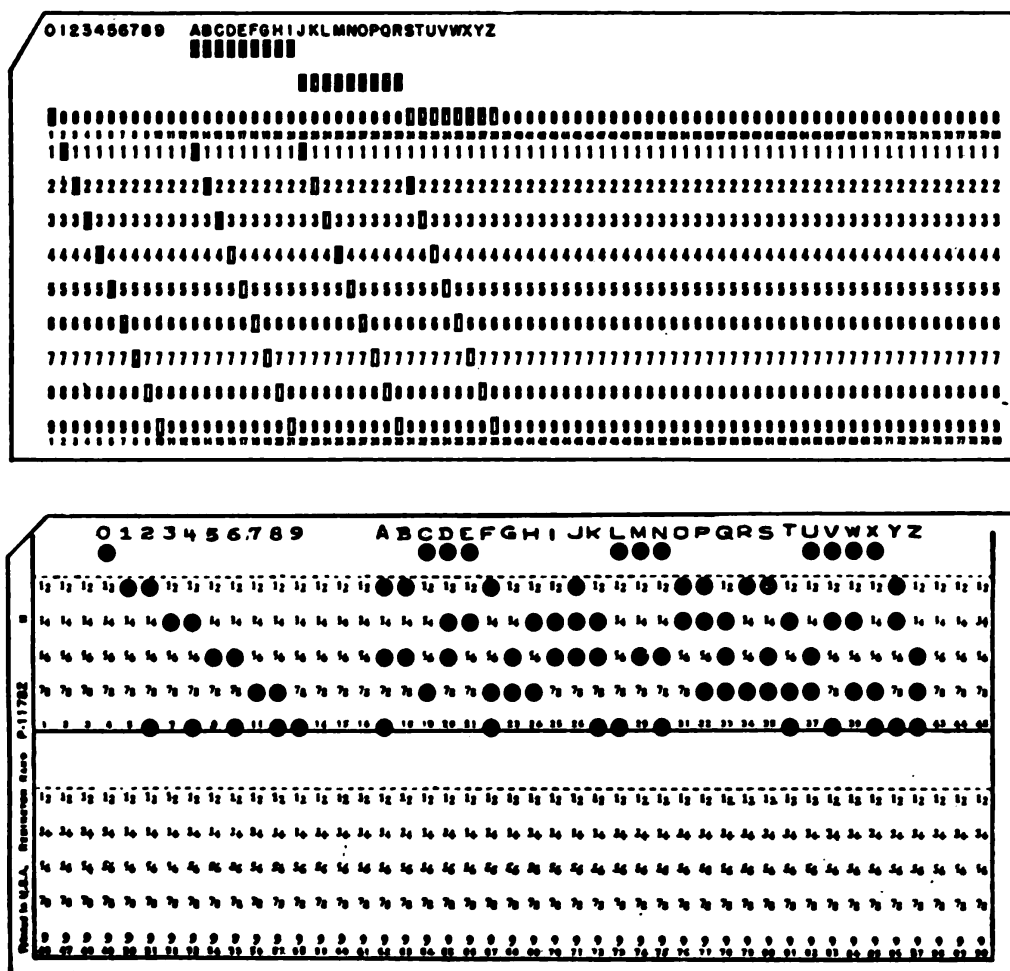


Figure 1-3. Machine-sorted punched cards. Symbols, coded in vertical columns, are printed at top of card.

can be selected from a file. Also, the file can be sorted into numerical or alphabetical order. The numerals or letters coded on the card can be printed on the card or on another form. Calculations can be made, coded cards reproduced, and numerous other operations carried out, all of which are controlled by the patterns of holes on the cards and by the use of the appropriate machines. Other systems, proposed or in various stages of development, use jets of air, light rays, or fluorescent, radioactive or magnetic spots and shapes for actuating the mechanisms.

Some of the Why of using punched cards in scientific information work has been summarized in many publications. Hill, Casey and Perry,³ in an article entitled, "Research and Chemical Information," say: "There is

³ Hill, Norman C., Casey, Robert S., and Perry, James W., *Chem. Eng. News*, 25, 970 (1947).

grave danger that libraries of chemical information may become mere warehouses of sheeted cellulose. The rapid and accelerating rate of increase in published scientific material can only render the situation more acute.

"The chemical industry is faced with much the same problem as the telephone companies when they realized years ago that the time was fast approaching when there would not be enough qualified switchboard operators to service the increasing number of calls. Just as the telephone industry developed the dial system, the chemical information field must develop newer, faster means and mechanisms for locating and correlating chemical information." The same reasoning might be applied to fields other than chemistry.

Frear,⁴ in "Punch Cards in Correlation Studies," says: "Experience gained in two years of work in this laboratory has led to the conclusion that punched cards are adaptable to a wide variety of chemical problems, particularly those which deal with large groups of data, such as are encountered in surveys, correlations or extended experimental studies.

"The particular problem we had here was a statistical investigation of the correlation between chemical structure and toxicity towards insects and fungi. From a search of the literature and other sources, data were collected on approximately 8,000 compounds, on each of which one or more toxicity tests had been made. With such a large number of compounds, and such a wide variety of constituent groups, counting and correlating the data by inspection promised to be a formidable task."

He then briefly describes his work in using hand-sorted cards and adds: "By slight modifications of these basic principles, it is possible to make correlation studies between chemical constitution and any desired property, chemical or physical."

Cox, Bailey and Casey⁵ reported their work with hand-sorted cards in "Punch Cards for a Chemical Bibliography": "In most of the bibliographic files which the authors have seen, the emphasis was on the manner in which the data are to be put *into* the file. The basis of the system described below is facility in getting desired data *out* of the file." They also state, "Crane and Patterson in a discussion of the preparation of bibliographies have said: 'As to arrangement, there are at least four possibilities: by dates, by authors, by sources, and by subjects.' With the proposed (punched-card) system, it is not necessary to choose one of the above categories for the arrangement of the references. One can arrange or segregate the cards according to any of those categories, and include other classes as needed, still using only one card for each reference."

⁴ Frear, Donald E. H., *Chem. Eng. News*, **23**, 2077 (1945).

⁵ Cox, Gerald J., Bailey, C. F., and Casey, Robert S., *Chem. Eng. News*, **23**, 1623 (1945).

TABLE 1-1.-LITERATURE SEARCHING MACHINES AND SYSTEMS

Name of Systems or Equipment	Examples	Manipulative	Storage		Delivery
			Indexing Possibilities	Number of Documents	
1. Hand-sorted punched cards	McBee cards, E-Z sort cards, Flexisort cards	Desired cards selected by manual manipulation of needles in holes or slots. Multiplicity of needles may be used with certain auxiliary devices.	<i>Direct coding</i> : limited by number of holes in cards (up to about 200) <i>Superimposed coding</i> : limited in number of index entries (2 to about 20) <i>Relationships</i> : not convenient to record complex relationships among index entries	Limited by tolerance of user to needling operation*	Actual card containing abstracts, indexes, and bibliographic information.
2. Hand-manipulated aspect cards (a) number matching (b) identification of pattern coincidence	(a) Uniform cards (b) Batten cards, Uniform cards	(a) Columns of numbers on one "aspect" card matched visually against numbers on another aspect card (b) "Aspect" cards superimposed, and desired document numbers detected visually by light passing through coincident holes	Limited by increasing number of false combinations as "depth of indexing" increases	Limited in convenient use by quantity of document numbers that can be recorded on a single "aspect card"	Document number is identified which then serves as entree to document file arranged in accession number order.
3. Machine-sorted punched cards (a) Fixed field (b) Intermediate (c) "Free field"	(a) Remington Rand, IBM, and Powers Series "standard" sorters and collators (b) IBM "101" (c) IBM Luhn Scanner, experimental IBM X 794, Census "100"	Punched cards maintained in drawers and fed into machine in stacks of up to 500.	(a) Similar to hand-sorted punched cards, above; certain relationships may be shown (b) Similar to hand-sorted punched cards, above; certain relationships may be recorded (c) Unlimited number of index entries per document; certain relationships among index entries may be recorded	(a) Limited by tolerance of user to sorting operations in fixed fields* (b) More convenient than (a)* (c) More convenient than (a) and (b) for certain applications*	Document number is identified which serves as entree to document file. Limited amount of bibliographic material or data may be printed on face of card or recorded on microfilm insert

INTRODUCTION

9

4. Machine-manipulated aspect cards (a) number matching (b) pattern coincidence	(a) Experimental machines developed on ONR funds (b) Peek-a-boo (NBS)	Same as for 2 (b) above, except that coincidence numbers are detected either mechanically or by activation of light sensitive terminal.	Same as for hand-manipulated aspect cards	Same as for hand-manipulated aspect cards
5. Tape-reading devices (a) magnetic tape (b) punched paper tape	(a) IBM "700" series, Remington Rand Univac series digital computer (b) WRU Selector	Tapes are stored on reels, entries are scanned either electronically or by means of mechanical "feeders." Entries may be transferred automatically to duplicate or output tapes.	Limited by time available for search as opposed to scope of search desired.*	Document number is identified which serves as entree to document file. Entire encoded "message" or abstract may be reproduced if desired.
6. Photographic film devices (a) Fixed field (b) Intermediate (c) Free field	(a) Rapid Selector (b) Filmorex (c) Minicard, Photoscopic storage device (International Telemeter)	(a) Film is stored on reels, and is transferred to other reels as scanned (b) Unit pieces of film manipulated and scanned (c) Encoded information on film is scanned in sequence, stored on unit pieces of film or rolls, or discs	Same as 5, above.	Selected film or film card contains coded information, as well as microcopy of document. For photoscopic storage device, same as 5, above.

* Subfiles or total files may be searched as desired.

Reprinted from: A. Kent, *Am. Doc.*, 8: No. 2, 150-151 (April 1957).

In "Some Applications of Punched-Card Methods in Research Problems in Chemical Physics," King⁶ summarizes his discussion of machine-sorted cards as follows: "The use of punched cards in research problems can be divided into three types: (1) large-scale repetition of simple operations, such as addition, subtraction and multiplication, modified if necessary by elaborate classification and selection; (2) a feasible stochastic or trial-and-error approach to the solution of problems; and (3) the construction of a representative sample of a population for statistical analysis.

"These principles are illustrated by their use in preparing tables of thermodynamic functions of compounds and spectrum analysis, and the calculation of the configuration entropy of high polymers."

Thus, the fundamental reason for using punched cards is that their use facilitates many routine and repetitive operations involved in the solution of certain intellectual problems. This is particularly true of problems in which large masses of data are involved. The machines can do some things which, due to their complexity and the amount of labor involved, could hardly be undertaken otherwise.

Each new milestone in human progress presents new problems as well as opportunities. For example, the widespread use of printing as a means for recording and disseminating scientific information has been most beneficial. However, as Vannevar Bush⁷ has pointed out, "Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it; and this sort of catastrophe is undoubtedly being repeated all about us, as truly significant attainments become lost in the mass of the inconsequential. The difficulty," says Dr. Bush, "seems to me, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record."

The punched-card technique is opening up new possibilities for coping with the growing mountain of research publication. These possibilities, however, are accompanied by problems, some of which are immediately apparent. One of the principal problems is the need to provide more precise methods for analysis and organization of information.

Punched cards have aroused the interest of index and classification experts, whose activities have of necessity been limited by their tools, namely, a set of pigeon holes or its equivalent for classification, and alphabetized lists of words either on bound sheets or in conventional card files for indexing. The introduction of the punched-card technique has broadened the horizons of indexing and classifying and has opened new territory which is now being cultivated.

⁶ King, Gilbert W., *J. Chem. Ed.*, **24**, 61 (1947).

⁷ Bush, V., *Atlantic Monthly*, **176**, 101-8 (1945).

The extent to which the field has been cultivated is evident from the many publications discussing research, development, and applications of new devices, tools and systems which are recorded in the Bibliography (Chapter 30).

The newer tools and systems for literature searching have been tentatively classified recently by Kent and Geer⁸ in Table 1-1.

Punched cards and related devices cannot make all complex problems simple, but they can make some complex problems less complex and some simple problems less tedious and time-consuming.

⁸ Allen Kent and Harriet Geer, "Searching the Chemical Literature Mechanically," Paper presented before the American Chemical Society, Sept. 1956, Atlantic City, N. J. Table reprinted from A. Kent, *Am. Doc.*, 8, No. 2, 150-151 (1957).

•

Chapter 2

ELEMENTARY MANIPULATIONS OF HAND-SORTED PUNCHED CARDS

ROBERT S. CASEY

W. A. Sheaffer Pen Co., Fort Madison, Iowa

AND

JAMES W. PERRY

*Center for Documentation and Communication Research, Western
Reserve University, Cleveland, Ohio*

The How of using punched cards has two aspects. One of these is mechanical and involves learning how to punch, sort and otherwise manipulate the cards. The other aspect is intellectual and concerns the necessity of analyzing the subject matter to which punched cards are to be applied. One must determine what meanings are to be assigned to the holes in the cards and what mechanical manipulation of the data is required, as discussed briefly under "Subject Analysis and Coding" later in this chapter. Although it is not possible to discuss either of these aspects apart from the other, this chapter is concerned principally with the first aspect—the basic mechanics of punched-card techniques.

Description of Cards

Hand-sorted punched cards may be obtained commercially in a variety of sizes, from $1\frac{1}{2} \times 2\frac{1}{2}$ up to $8 \times 10\frac{1}{2}$ inches. One widely used type ("Keysort") has one or two rows of $\frac{1}{8}$ -inch holes on $\frac{1}{4}$ -inch centers parallel to the edges of the cards, with the first row $\frac{1}{16}$ inch from the edge. Another type ("E-Z Sort") has elliptical holes spaced 6 to the inch. With all types of hand-sorted, edge-punched cards, the holes along the edges occupy only a small fraction of the total card area. Consequently, most of the area on both sides of the card is available for writing, typing or printing references, abstracts, observations, and numerical data, as well as for coding directions and attaching pictures, clippings and other thin, flat material.

One corner of each card is cut off so that it is possible to see at a glance that all the cards are right side up and facing the same way. The holes in the other three corners are never punched. They are used to arrange the cards right side up and to see that they face forward should they become mixed (*see page 18*).

BASIC SORTING OPERATION

Direct Sort of Outer Row Holes

The basic sorting operation separates the cards punched to form a notch in a given position from those not so punched. As shown in Figure 2-1, insertion of the tumbler or sorting needle into the hole in question, and then raising the tumbler or needle, permits the cards punched (notched) in that position to fall, unless they are prevented from doing so by the friction of other cards. In order to facilitate the dropping of cards punched in the position being sorted, the following technique has been developed.

Figure 2-1. Remove from the file a group of cards not more than 2 inches thick and place them in a vertical position on the alignment block with the hole portion to be sorted at the top. Jog the cards against the vertical edge of the alignment block to align the holes. Support the cards with the left hand. Place the left thumb adjacent to the hole to be sorted and compress the cards with the thumb and fingers of the left hand. Grasp the handle of the tumbler firmly in the right hand, palm underneath the handle. *Keep the tumbler horizontal at all times* to prevent the cards falling off the needle or sliding back against the handle. Insert the needle into the hole to be sorted, guiding it with the left thumb. Push the needle through the cards, leaving at least one inch between the tumbler handle and the front card.

Note—The alignment block, shown under the cards in Figures 2-1 to 2-6 inclusive, is a sheet metal device which fits against the front edge of the desk. A portion at the right edge is bent into a vertical position perpendicu-

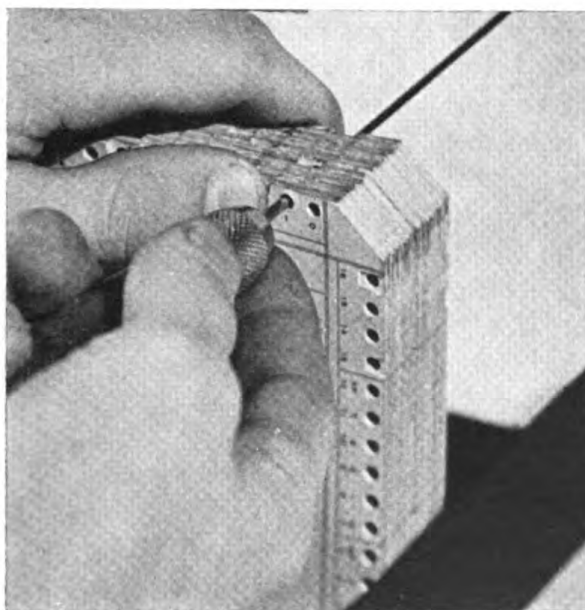


Figure 2-1. Start of single needle direct sort.

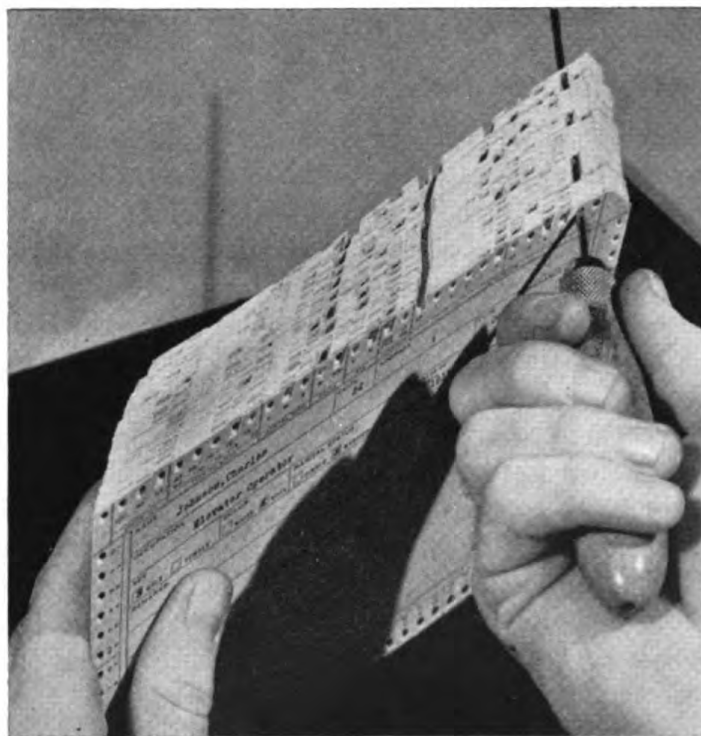


Figure 2-2. Second step in sorting cards.

lar to the front edge of the desk. This device is not absolutely necessary, but assists in sorting the cards as described below.

Figure 2-2. Support the cards loosely with the left hand at the left vertical edge of the cards. Swing the handle of the tumbler toward the left, pushing the cards toward the right with your left hand. This cramps the cards in a diagonal fashion as shown in Figure 2-2. While the cards are in this position, grasp them firmly between the thumb and fingers of the left hand.

Figure 2-3. While holding the left edge of the cards firmly with the left hand, move the handle of the tumbler toward the right, back to its original position. This causes the cards to fan out and separate as shown in the illustration. Hold the cards in this position and with both hands lift them several inches above the alignment block. *Keep tumbler horizontal.*

Figure 2-4. Release the left-hand grip on the cards, at the same time giving them a slight downward jerk with both hands. Hold the left hand as shown in the diagram, forming a U, and lightly support the cards which drop. *Do not grasp falling cards tightly.* If the cards are compressed at this point, it will hinder their separation. Swing the sorting needle back and forth gently from left to right so that the suspended cards pivot about the needle. This motion facilitates dropping the cards being sorted.

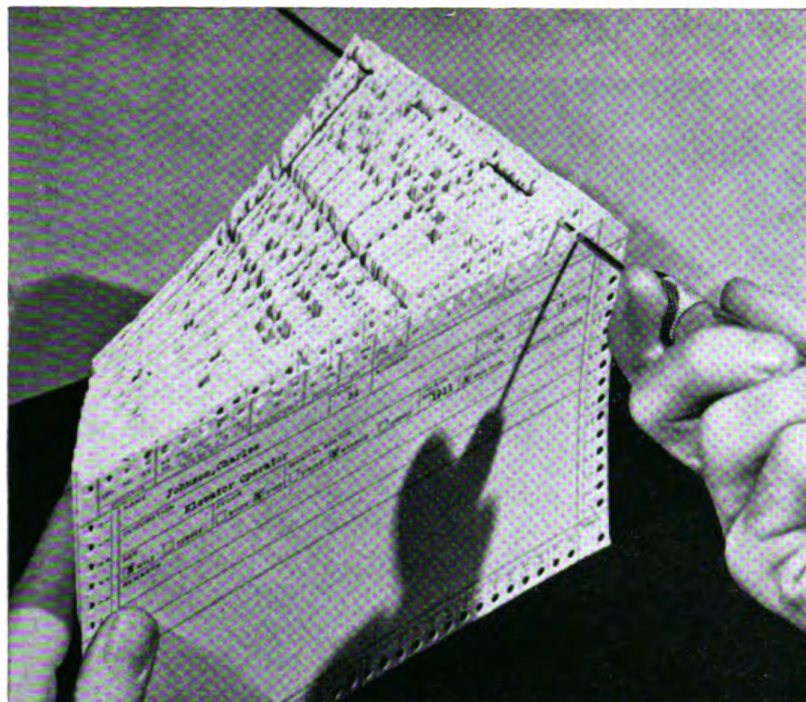


Figure 2-3. Third step—cards are fanned out.

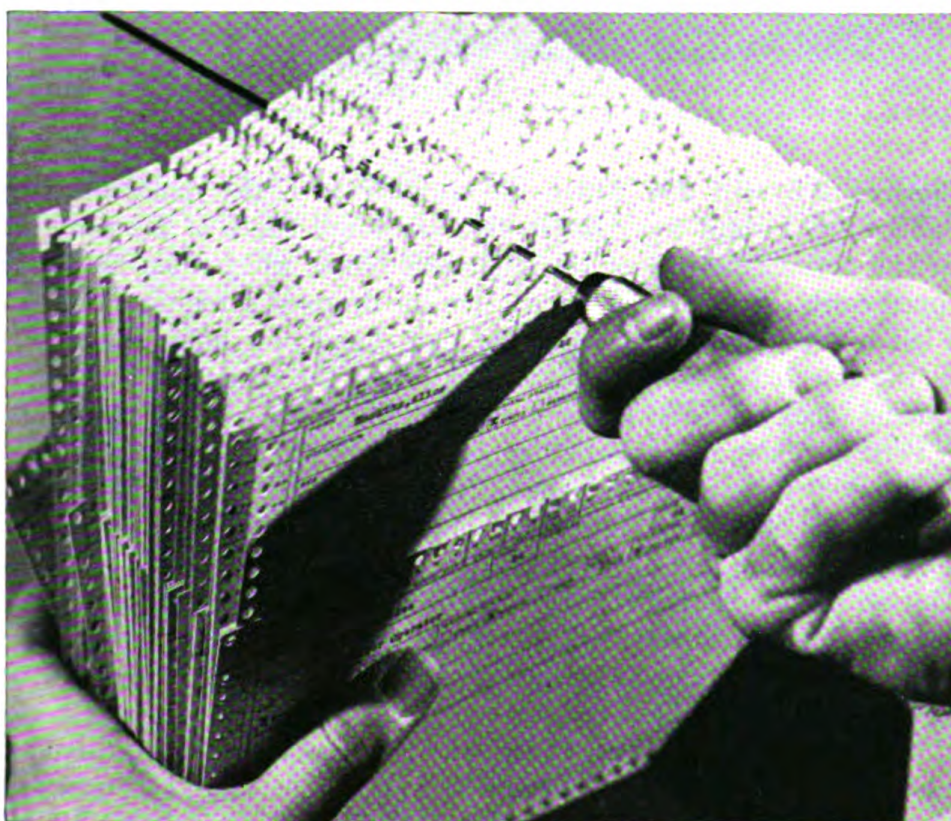


Figure 2-4. Fourth step—separation of cards begin.

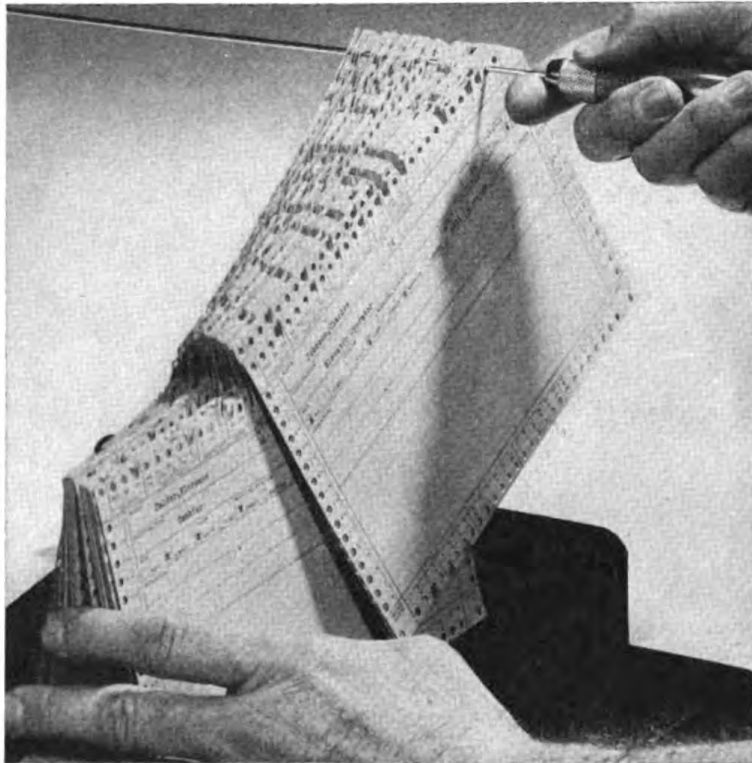


Figure 2-5. Fifth step—vertical edge on alignment block helps separate dropped cards.

Figures 2-5 and 2-6. Move the tumbler with the remaining cards hanging on it toward the right so that the lower edges of the cards just clear the vertical portion of the alignment block, which then retains the dropped cards. With the left hand place the dropped cards on the desk at the left of the alignment block. The right hand now holds the tumbler with the rejected cards still suspended. Place the rejected cards on the alignment block before removing them from the tumbler. Twist the tumbler in a vertical plane, moving the handle downward. This spreads the cards so each one is higher than the one in front of it. Scan the top of the cards just sorted to see if any cards edge-punched in that hole failed to drop. If so, remove them and place them with the selected cards.

The series of operations just described can be completed in about 15 seconds or less by a person who has had only a moderate amount of experience. The operations are then repeated on additional groups of cards until the complete file has been sorted.

If the position to be sorted is near the left edge of the card, there may not be enough space between the tumbler and the left hand to spread the cards as illustrated in Figure 2-3. In this case, move the left hand to the bottom of the cards directly below the tumbler and twist the tumbler in a

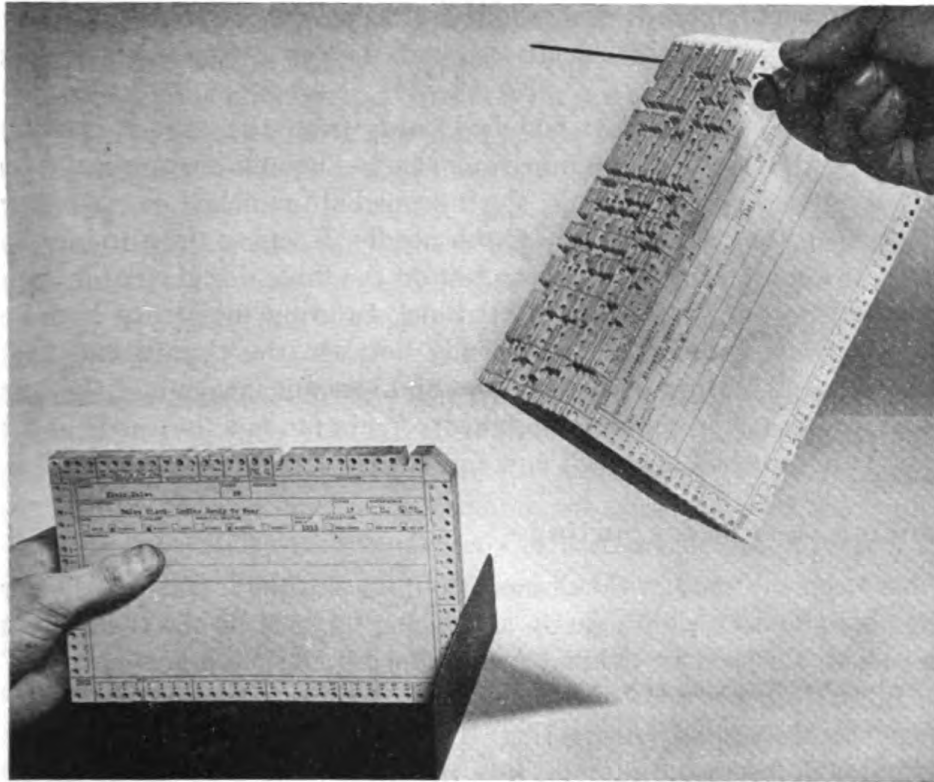


Figure 2-6. Sixth step—separation of selected from rejected cards is completed.

vertical plane, moving the handle down. Grasp the cards at the bottom edge firmly between the thumb and fingers of the left hand and twist the tumbler back to its original horizontal position. This gives the same result as shown in Figure 2-3.

Sorting Operations Involving Double-Row Holes

Double-row holes may be punched in three different ways as shown in Figure 2-7. Cards deep-punched in a given position may be separated from all the others by inserting the needle in the inner hole and carrying out the sorting operation described for sorting outer row holes. Similarly,

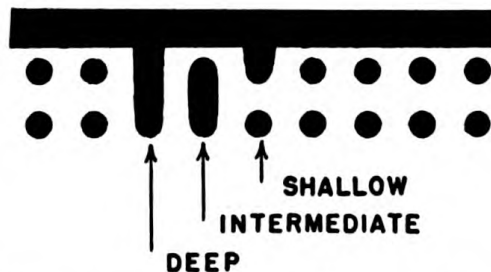


Figure 2-7. Three different ways of punching each position in a double row of holes.

separating deep-punched and shallow-punched cards from cards not punched at all, or only intermediate-punched, is exactly the same as sorting single-row holes.

Separation of intermediate-punched cards from the others is effected by inserting the needle in the inner hole, spreading and dropping the cards as already described. As a result, the intermediate-punched cards will drop about $\frac{1}{4}$ -inch, but will not fall off the needle. Keeping the tumbler horizontal, but not allowing the cards to rest on the table, jog the cards against the vertical portion of the alignment block to bring them into horizontal alignment. Now, grasp the cards firmly between the thumb and fingers of the left hand. Withdraw the tumbler and insert it into one of the corner holes which are never punched (notched). Again fan out the cards and the intermediate-punched ones will now fall clear.

Arranging the File for Sorting

It has been assumed in the discussion up to this point that all the cards in the file have been right side up and facing forward. If the cards are not so arranged, this will be evident from glancing at the corner cut-off. The holes in the other three corners, which are never punched, are used to arrange the file properly for sorting. If the upper right hand corner of the file is clipped, insert the tumbler into the hole showing at the upper right corner of the file, and proceed with the sorting operation as previously described. The cards which drop are right side up and facing forward. Repeat, after turning the cards which remain on the tumbler through 180° in their own plane. Then rotate the cards which still remain on the tumbler 180° perpendicular to their plane, and again needle the upper right hole.

ELEMENTARY SUBJECT ANALYSIS AND CODING

Specific meanings must be assigned to the holes with consideration for whatever subsequent sorting may be required. In general, this is done with one or both of two purposes in mind, namely, *selecting* certain cards from the file or *arranging* all the cards in a given order.

Coding is matching the idea, datum or concept with a punched or notched hole or a pattern of punches. For convenience, a set of symbols is interposed between concept and card. For example, the punching or notching positions on cards are usually numbered. Then the file is "coded" by making a list of subject headings and assigning one of the numbers or combinations of numbers or other symbols indicated on the card to each entry. The term "code" is also used to refer to the pattern of symbols on the card. Letters of the alphabet can be indicated by directly labeled positions or by numbers. Confusion may arise because on different occasions a letter or a number may be either a concept or a symbol.

Remember that the symbols are merely an intermediate convenience. The code is: a certain pattern of punches equals a certain concept.

In setting up a punched-card file one must analyze the purposes of the file and the ways in which the data on the cards are to be used. Thus, with a bibliographic file, one should be able to *select* the cards bearing references concerning any one of the subjects of interest to the bibliographer. Also, it may be desirable to *arrange* the whole file alphabetically by authors or chronologically by date of publication. If one wishes also to *select* cards according to author and date of publication, one must choose the appropriate type of code.

Direct Coding. In this, the simplest form of coding, a separate meaning is assigned to each hole. All the cards on which the meaning is coded (e.g., all references published in the nineteenth century, or those describing analytical procedures or chemical compounds which contain the hydroxyl group) are separated with a single pass of the sorting tumbler. The application of direct coding has its limitations. For example: There may not be enough holes on the card to code all of the desired data and more passes of the sorting tumbler are required for serial sorting than are required for numerical codes.

Numerical Codes. A numerical code might be called a "combination" code since one or more holes may be punched to represent a single number, letter or other entity. The most commonly used numerical code is illustrated in Figure 2-8.

By punching various combinations of the four holes marked, respectively, 7, 4, 2, and 1, one may code any number from zero (no punching) up to and including fourteen (all holes punched). Such a group of holes is called a "field." This code is a modification of the 1, 2, 4, 8, 16 . . . series; 7 is used instead of 8, so that with four positions any digit may be indicated by punching not more than two holes. By using one such field each for units, tens, hundreds, etc., relatively few holes are required to code any one large number that may be desired (Figure 2-9).

A file of cards coded in this way is arranged into numerical order by sorting in order from right to left each hole in the numerical fields and placing the cards which drop at the back of the pack before sorting the next hole. Detailed instructions for this sorting procedure are given later in this chapter.

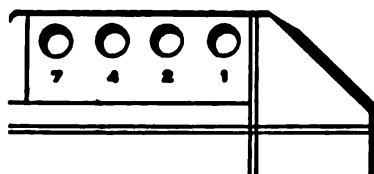
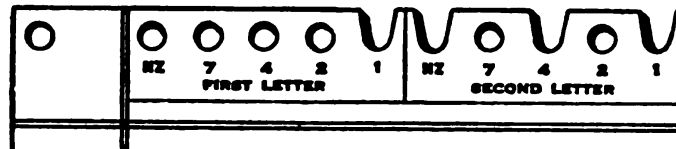


Figure 2-8. A numerical sorting field.



Figure 2-9. A 5-Digit numerical sorting field.

Figure 2-10. A 2-field, 5-hole alphabetic code.
Letters coded: A R.

Although large numbers can be coded in numerical fields, only one number can be coded in any one field on any one card. Thus, such coding is used for numbered lists of data which are mutually exclusive such as serial numbers. Although such coding provides a convenient means for sorting the file into serial order, unequivocal selection of a card coded for a given number is not possible. (See Selector and Superimposed codes, below)

Alphabetical Codes. Alphabetical codes, like numerical codes, are based on the use of combinations of holes, with the exception that the coding represents letters of the alphabet instead of numbers.

A commonly used alphabetical code which is, in fact, a variation of the numerical code, is illustrated in Figure 2-10. The letters A to M, inclusive, are numbered consecutively 1 to 13, and the appropriate number is coded to represent the desired letter. If the desired letter is in the second half of the alphabet, the letters N to Z are coded in the same manner, with the additional punching of the N-Z hole.

Alphabetical codes, as well as numerical codes, can be used to sort a file into serial order; in fact, the same general comments made concerning numerical codes also apply to alphabetical codes.

Chronological codes can be made up by adapting a direct or combination code to the need of the user. If it is necessary to code only the year, one may use two 7, 4, 2, 1 fields for the units and decades, respectively, and one or more other holes for the century, depending on how long an interval is to be covered.

Selector Codes. In some punched-card files it is desirable to be able to select all cards of a certain category. If there are more categories than holes available for direct coding, selector codes are necessary. Selector codes are special combination codes so conceived that, in the simplest case, two holes, no more and no less, are punched in each field to represent each symbol. Then, when two sorting needles are inserted into the holes representing the desired symbol and lifted, only the desired cards drop.



Figure 2-11. A 6-hole selective code, punched to indicate number 45.

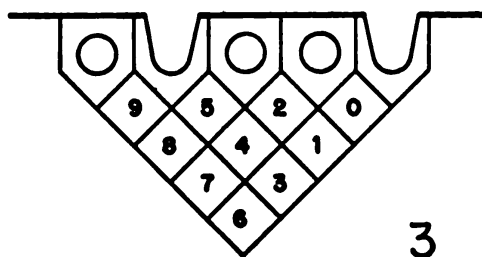


Figure 2-12. A 5-hole triangle or pyramid selective code, with symbols arranged in proper order to permit serial sorting also.¹ Number coded 3. See also Fig. 21-1.

A commonly used selector code has the positions 0 (zero) and SF (single figure), in addition to 7, 4, 2, 1. To code 1, 2, 4, or 7 the SF hole is punched in addition to the numbered hole. The other digits require the punching of two of the numbered holes. To code zero the 0 position alone is punched (Figure 2-11).

Another simple selector code is the triangle or pyramid type shown in Figure 2-12. It is coded by punching the holes at the tops of the mutually perpendicular diagonal columns which intersect at the desired symbol. In the example shown, the two columns are those containing the numbers 0, 1, 3, 6 and 5, 4, 3, respectively.

These selector codes permit only one number to be coded in common with the other combination codes described previously.

Double-Row Coding. The codes described above are for a single row of holes. A double row of holes increases the coding possibilities in a given amount of card space since each position in a double row may be punched in any one of three ways, using the shallow, intermediate, or deep punch, (Figure 2-7). The inner and outer rows of holes can also be coded and sorted independently, by punching the outer row of holes with the shallow punch, and the inner row with the intermediate punch, which clips open the space between the inner and outer hole.

Superimposed Coding. The direct and combination coding schemes already described are often inadequate to code the number of subject concepts that are required. Superimposed coding permits one to code on each card several concepts selected from a list of a great many more concepts than there are holes in the card. Each concept is coded by notching two or more of the positions along one edge of the card.

¹ Cox, Gerald J., Robert S. Casey and C. F. Bailey, *J. Chem. Ed.*, 24, 65 (1947).

The system can be designed with the combinations assigned randomly so that the chance of getting unwanted cards with a given sort is negligible. For example, if one sorts for "disease," which is assigned holes numbered 7 and 13, and for "antibiotic," 4 and 27, and for "animal," 15 and 19, one would get only a negligible number of "extra" cards coded 4 and 13, 7 and 19, 15, and 27.

Various modifications and applications of the principle of superimposed coding are described in Chapters 10, 15, 18, 21 and 23, and in various references in the bibliography.

Subject Analysis. Some sort of analysis of subject matter is necessary no matter what kind of file or system is to be set up. A collection of reprints or a file of references on plain cards may be filed in alphabetical order by author, grouped according to subdivisions of the main subject, or indexed in a notebook. Often additional cross-reference cards are made to cover different subject aspects of a given reference. Books and pamphlets are placed on shelves in some sort of order.

When setting up a punched-card file, however, *special* consideration must be given to subject analysis in order to make full use of the advantages offered by punched cards (Chapter 1). New considerations are necessary, or rather a new viewpoint concerning the extensions and combinations of older considerations is necessary.

One of the most valuable properties of punched cards is their multi-dimensional or multi-aspect coding possibilities. Each of various independent aspects of the subject matter can be coded independently. For example, substances can be listed according to their composition or form, together with their chemical or physical properties or their functions such as "solvent," "antioxidant," "lubricant," and "fungicide." Another aspect is "processes and procedures," such as "analyze" or "measure," "oxidize," "distill." Others are "conditions," "energy manifestations," and "structures" such as "machines" and "apparatus."

Then several broad entries from various aspects can be combined to define a specific bit of information. A thermometer is "apparatus" to "measure" "thermal condition." One can select cards bearing information about "antibiotics" in the "treatment" or "therapy" of "gastrointestinal" "diseases" in "animals."

Avoid making the entries under each aspect too specific. Make the entries as broad and generic as is consistent with the subject matter in the file. The coding should be done only after careful study of the references. It may be necessary to make a generic search, say for "antibiotics." If antibiotics have been coded individually, sorting for "penicillin," "streptomycin," and the others one by one would be required. Furthermore, if one is seeking references on an individual antibiotic, it may be easier to go through several dozen cards coded "antibiotics" by hand, and pick out

the relevant ones than it would be to make a more elaborate and specific sort.

Don't try to set up a classification scheme or schedule of subject entries *a priori*. Study the references. Think of the *ideas* which are discussed, not the words used to describe the ideas. Ask yourself, "What are some of the questions this document would answer?" "Why would I ever want to find this document?"

Think generically. Don't code individually "children," "men," "women," "Caucasians," "Mongolian idiots" when you can use "humans" and not cover too many references. Start with "humans," and leave some of the holes in the cards unassigned so you can subdivide later, if necessary. For some files "humans" might be too specific, "animals" or "living organisms" could be used.

Each subject entry should cover a carefully defined area of meaning. The entry does not need to be a single word, it can be a phrase, a sentence, a paragraph, or a diagram of a structure. The entry can be any notion the human mind may conceive, but its area of meaning should be carefully defined.

Subject analysis and coding are discussed in more detail in Chapters 18, 24, and 25 and in most of the chapters in Part II. Coding of chemical compounds is described in Chapter 22. The mathematical analysis of coding is developed in Chapter 21.

SPECIAL SORTING TECHNIQUES

As already pointed out in discussing "Basic Sorting Operations," it is a very simple process to separate cards punched (notched) in a given position from those not so punched. By combining a succession of simple sorting operations with certain special codes, it is possible to effect the arrangement of a file into a numerical or alphabetic sequence. The technique of doing this is discussed in the following paragraphs.

Sequence Sorting. Perform the sorting operations illustrated in Figures 2-1 through 2-6 on each hole and in order from right to left through the fields composing the numerical or alphabetical code. The cards which drop after each sort are placed in back of the cards which remain on the tumbler before the next hole is sorted. All the cards which drop must be kept in their relative order. All cards which fall out of position, or fail to drop from the group hanging on the tumbler, must be laid aside and filed manually after the sorting is completed.

After each sort, jog the dropped cards into alignment with the left hand. Then, with the right hand still grasping the tumbler handle, return the rejected cards to the alignment block in front of the dropped cards. Glance along the edges of the cards remaining on the tumbler at the hole just sorted to see if any cards failed to drop. Next bring the dropped and re-

jected cards into alignment, allowing the needle to slip into the groove in the group of cards just dropped. Grasp the cards firmly between the thumb and fingers of the left hand, with the thumb adjacent to the left edge of the next hole to be sorted. Remove the tumbler from the hole just sorted and insert it into the next hole to the left, repeating the sorting operation just described. After sorting the last hole at the left end of the coded portion, the cards will be in serial order. The above process is called *fine sorting*.

If there are more cards to be sequence sorted than can be handled on the tumbler at one time, the above fine sorting procedure must be preceded by a breakdown sort. In general, the breakdown sort is carried out by sorting each hole from left to right and by placing in separate piles the cards which drop when each hole is sorted until the whole file is arranged in piles small enough to be fine sorted. The following example will illustrate the procedure:

9,999 cards, numbered from 1 to 9,999 and coded by the 7, 4, 2, 1 system, are to be arranged in numerical order. Take from the file a group of cards not over two inches thick.

(1) Sort through the 7 hole in the "thousands" field. 7xxx, * 8xxx, 9xxx cards drop and are placed at the left rear portion of the desk as the first pile.

(2) Sort the 4 hole in the "thousands" field. 4xxx, 5xxx, 6xxx cards drop and are placed in the second pile at the right of the first pile.

(3) Sort the 2 hole in the "thousands" field. 2xxx, 3xxx cards drop and are placed in the third pile at the right of the second.

(4) Sort the 1 hole in the "thousands" field. 1xxx cards drop and are placed in the fourth pile at the right of the third. Cards numbered 1 to 999 remain on the needle and are placed in the fifth pile at the right of the fourth.

(5) Repeat operations (1) to (4) with successive groups of cards until all cards have been distributed to the five piles.

(6) Repeat operations (1) to (4) in the "hundreds" field on the fifth pile, which contains cards numbered under 1000. This makes additional piles as follows:

7xx-8xx-9xx

4xx-5xx-6xx

2xx-3xx

1xx

1 to 99

* "x" indicates any digit, 0 to 9.

(7) Fine sort the 1 to 99 group and place it in the front of the file storage drawer. Fine sort the 1xx group and place it in the file behind the 1 to 99 group. As each group is fine sorted, put it in the file in this order.

(8) Sort the 1 hole in the "hundreds" field in the 2xx-3xx group. Fine sort the 2xx group, which remains on the needle, and place it in the file. Fine sort the 3xx cards, which have dropped, and place them in the file.

(9) Sort the 2 hole and then the 1 hole in the "hundreds" field in the 4xx-5xx-6xx group. Fine sort each of the three resulting piles in this order: the 4xx cards, which remained on the needle; the 5xx cards, which dropped when the 1 hole was sorted; and the 6xx cards, which dropped when the 2 hole was sorted. Then place them in the file.

(10) Sort the 2 hole and the 1 hole in the "hundreds" field in the 7xx-8xx-9xx group. In the same order as described under (9) fine sort the three resulting piles and place them in the file.

(11) Repeat operations (6) to (10) with the fourth pile, which contains cards numbered 1000 to 1999.

(12) In their turn, sort the other piles, first into thousands and then into hundreds. Then fine sort and file each hundred as outlined in the preceding steps.

If two persons are sorting, the breakdown sorting should be done by one person and the fine sorting by the other.

Multiple Sequence Sorting. A punched-card file may be arranged in serial order according to one category ("Major" item), and, at the same time, in serial order according to a second category ("Minor" item), under each item of the first.

Multiple sequence sorting is used in correlation studies, such as the physiological properties *versus* chemical structure discussed by Frear in Chapter 22. It may also be used for arranging a bibliographic file in chronological order according to the date of publication of each subject covered by the file. To accomplish such an arrangement it is necessary to sequence sort the "Minor" items first, then sequence sort the file according to the "Major" item. This is clarified by remembering that we are doing the same thing when we sort several fields of a numerical code into serial order. The "units" are arranged first in serial order, then the "tens," and so on, until the "units" are in order under each digit of the "tens," the "tens" are in order under each digit of the "hundreds," and so on through as many places as there are in the code. This allows more than two categories to be put into serial order, one within the other.

If there are more cards to be sorted than can be handled conveniently at one time, breakdown sort the "Major" item. When that is complete, fine sort according to the "Minor" item. This becomes clear if one considers the codes of the "Major" and various "Minor" items as one continuous numerical code with each item, in order, occupying one decimal place.

Sorting Selective Codes. For sorting selective codes additional needles (without handles) and a tumbler are required. The needles used may be either metal rods or metal or plastic knitting needles, 2 millimeters in diameter and about 10 inches long.

In the selector codes two needles are inserted in the appropriate holes in each field (except for 0 in the Figure 2-11 code) to select cards punched for the desired number, letter or coded symbol.

Insert the tumbler into one of the holes required (if possible, near the center of the edge of the card for balance). Insert the loose needles in the other required holes.

Pivot the tumbler in the vertical plane, pressing the handle downward. Firmly grasp between the thumb and fingers of the left hand the bottom edge of the cards directly below the tumbler. Pivot the tumbler back to horizontal. Raise the cards a few inches and release the left hand, dropping the punched (notched) cards as described under "Direct Sort of Outer Row of Holes." This process is repeated with additional groups of cards, until the whole file has been sorted.

A group of cards not more than $1\frac{1}{2}$ inches thick should be taken for each selective sort. However, the number of cards which can be sorted conveniently at one time will be determined by experience.

This method of selective sorting requires that the needles be reinserted individually in each group of cards sorted. When the punched-card file is large and considerable selecting is to be done, it may be advantageous to use a selector unit described in Chapter 3. This unit has a metal bar provided with openings spaced the same as the holes along the edge of the card. The ends of the sorting needles are inserted into those openings which will space the needles suitably for the sort being contemplated. Thus, any number of groups of cards may be sorted without resetting the individual needles.

Most selector codes may also be serially sorted. The code illustrated in Figure 2-11 can be serially sorted by ignoring the 0 and SF positions and sorting the 7-4-2-1 fields exactly as described above for numerical codes.

The code illustrated in Figure 2-12 can be fine sorted serially by sorting each hole in turn from right to left as described previously for the 7-4-2-1 code, except that it is not necessary to sort the hole at the right. The fine sort can be started at the second hole from the right. The breakdown sort

is performed in a similar manner to that described for numerical codes. The holes are sorted from left to right, and the cards which drop as each hole is sorted accumulate into separate piles. Superimposed codes are sorted the same as selector codes.

SUPPLEMENTARY NOTES ON HAND-SORTED CARDS

Punched cards may be obtained with special forms printed on the card and the coding printed adjacent to the holes. Manufacturers also have certain standard cards which are more economical than specially printed ones. When standard cards are used, the coding and other matter may be overprinted or multigraphed, or the cards may be punched and sorted by referring to a master code card.

If the punching is to be done with a hand punch, the holes to be punched should be marked first. Avoid using soft crayon for this purpose because pieces of the crayon scrape off at the edges of the hole or card and cause unsightly streaks on the cards.

Precaution must be taken to prevent the cards from sticking together. To this end, avoid permanently tacky adhesives when attaching material to the cards. Paper, illustrations and samples of thin material may be attached to the cards by non-tacky plastic-base or heat-sealing adhesives. Even the buckling caused by water-base adhesives may not prohibit their occasional use. Do not put rubber bands around the cards. Rubber ages rapidly under tension and will leave tacky spots on the cards.

Bent or wrinkled cards will generally sort as well as, or better than, new cards. However, avoid adjacent cards becoming deformed in the same pattern. If the file becomes cramped and the corners or edges of the cards are bent so the cards tend to "nest," sorting will be hindered.

If these measures are heeded, a punched-card file may be stored and handled the same as an ordinary card file.

The information given in this chapter should be sufficient to enable an individual to perform the punching and sorting operations required to set up and use a simple file of hand-sorted punched cards.

An understanding of the mechanical features of punched cards is, however, only the first step toward their successful use as tools for handling information. Applying punched cards successfully to a problem also requires careful analysis of the subject matter involved, and its segregation and use. The chapters mentioned at the end of the elementary "Subject Analysis" section above, should be consulted.

It should be pointed out, however, that long and detailed study is not necessary before a person can obtain good results with a punched-card file. The beginner should strive for simplicity in coding. He should avoid complicated codes and too fine a breakdown of subject matter. It should

be remembered that it takes only a few minutes to glance over the material entered on several dozen cards. For this reason a code can be regarded as quite satisfactory if it permits sorting operations to select *all* the desired cards, even though they may be accompanied by a moderate number of undesired cards.

It is advisable at the beginning not to attempt to set up the code in final form. It is usually better to start out with a skeleton code built up of broad topics whose usefulness for effecting sorting operations cannot be doubted. This approach makes it possible to keep many holes in reserve for such finer analysis of the subject matter as may prove necessary.

Although a punched-card file can be kept in random order and sorted when necessary, it is often convenient to keep the file arranged in some order or segregated into broad subject categories.

Also it is convenient to assign a serial number to each reprint or other document as it is acquired, and to code, or at least write the number on the punched card. The documents can be kept in serial order and the cards in alphabetical order by author. Then a reference can be located by author or by serial number.

Edge-marking with tabs² and with pen or pencil^{3, 4} and edge-notching^{5, 6, 7} of plain cards have been suggested. The edges of a pack of plain cards can be fanned out and then scanned for a colored mark or notch in a given position on a given edge. This simple technique of broad subject searching might be tried if one has a file of plain cards, preliminary to adoption of punched cards.

Suggestions have been made for making one's own punched cards. A template and hand punch,⁸ a drill press,⁹ a specially constructed jig,¹⁰ and

² Reumuth, H., "The Indexing of Chemical Compounds. A Contribution to the Problem of Organization of the Literature," *Z. Angew. Chem.*, **41**, 1204-7 (1928).

³ Thurstone, L. L., "The Edge-Marking Method of Analyzing Data," *J. Am. Statistical Assoc.*, **43**, 451-62 (1948).

⁴ Lester, A. M., "The Edge Marking of Statistical Cards," *J. Am. Statistical Assoc.*, **44**, 293-4 (1949).

⁵ Aldrous, J. G., "Simple Method for Cross-Indexing a Reference File," *Science*, **106**, 109 (1947).

⁶ Campbell, D. J., "The Use of Notches in Cards as a Means of Signalling Information," *J. Documentation*, **9**, 224-5 (1953).

⁷ Schlink, F. J., "Getting the Most out of Index Cards," *Industrial Management*, **55**, 135-8 (Feb. 1918).

⁸ Begun, George M., "Making Your Own Punched Cards," *J. Chem. Educ.*, **32**, 328 (1955).

⁹ Thomas, George R., "The Preparation of Punched Cards for Indexing Information," *J. Chem. Educ.*, **29**, 406 (1952).

¹⁰ Thomas, Carl O., "A Jig for Preparing Edge Punched Cards," *J. Chem. Educ.*, **34**, 241 (1957).

a punch for preparing papers for plastic binders,¹¹ have been used to perforate the edges of either new cards or the cards in an existing file.

By following the simple precepts suggested in this chapter the beginner will find it possible to use his own experience as a basis for improving his skill in using punched cards.

¹¹ Cullman, Ralph E., (Letter to the Editor), *J. Chem. Educ.*, 30, 246 (1953).

Chapter 3

COMMERCIALLY AVAILABLE EQUIPMENT AND SUPPLIES*

THOMAS H. REES, JR.

*Center for Documentation and Communication Research, Western
Reserve University, Cleveland, Ohio*

This chapter is divided into five parts. The first deals with edge-notched punched cards and related equipment; the second with tabulating type punched cards; the third is concerned with supplements to punched cards; the fourth with other types of equipment related to information processing; and the fifth with ancillary equipment for punched-card systems.

All of the material presented here has been reviewed by the manufacturers or suppliers mentioned, and in most cases it includes descriptions taken from their literature. Replies from several commercial organizations were not received in time for inclusion and therefore descriptions of their products have been omitted.

EDGE-NOTCHED PUNCHED CARDS

The Keysort System

ROYAL MCBEE CORPORATION

Port Chester, N. Y.

The Keysort systems employ the marginal hole punched-card principle described in Chapters 1 and 2. All punching and sorting machines are manually operated, and their operation can be learned in a few minutes. Reasonable operating speeds are approached in a few days and usually attained within a week.

Royal McBee maintains representatives in more than 70 cities of the United States and Canada to service their systems and related equipment. Key punches, batch grooving machines, and card counters are available on a rental basis only, whereas all other equipment and supplies listed below may be purchased. No special wiring or building construction is necessary.

* Submitted in partial fulfillment of the requirements for the degree of Master of Science in Library Science, Western Reserve University, Cleveland, Ohio.

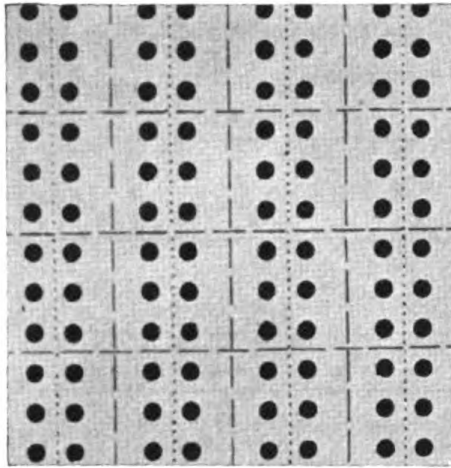


Figure 3-2. Keysort card savers.

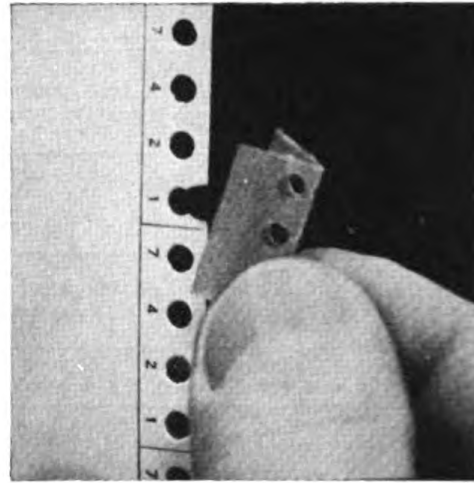


Figure 3-3. Keysort card saver in use.

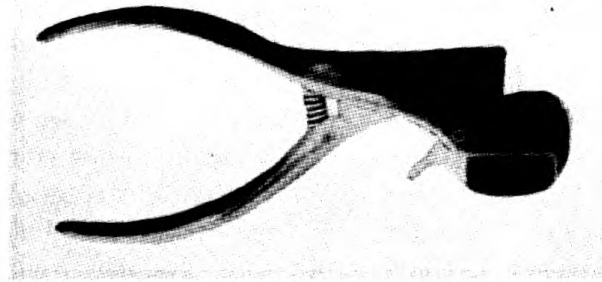


Figure 3-4. Keysort hand punch.

Keysort hand punches (Figure 3-4) are used for punching the cards. Three styles are available: shallow punch for single-row coding; deep punch for double-row coding (if card is inserted only halfway into throat in punch head, outside row is shallow-punched); and intermediate punch for inside double-row coding. (Intermediate punching is described in Chapter 2, page 17)

Keysorters (Figure 3-5) are used to make a direct sort, or to sort cards into sequence. Two types are available: Keysorter-manual (single needle—adjustable extension); and Keysorter-manual (single needle—fixed extension—supported at both ends)—recommended for sequence and single-needle selective sorting of cards $7\frac{1}{2} \times 8\frac{1}{2}$ inches and larger.

The speed of sorting will depend somewhat upon the operator. The manufacturer states that 60,000 single hole sorts per hour are considered average, although 90,000 have frequently been reported.

The **alignment block** (Figure 3-6) increases the speed and ease of sorting. The drop front fits flush against the front of the desk, which places

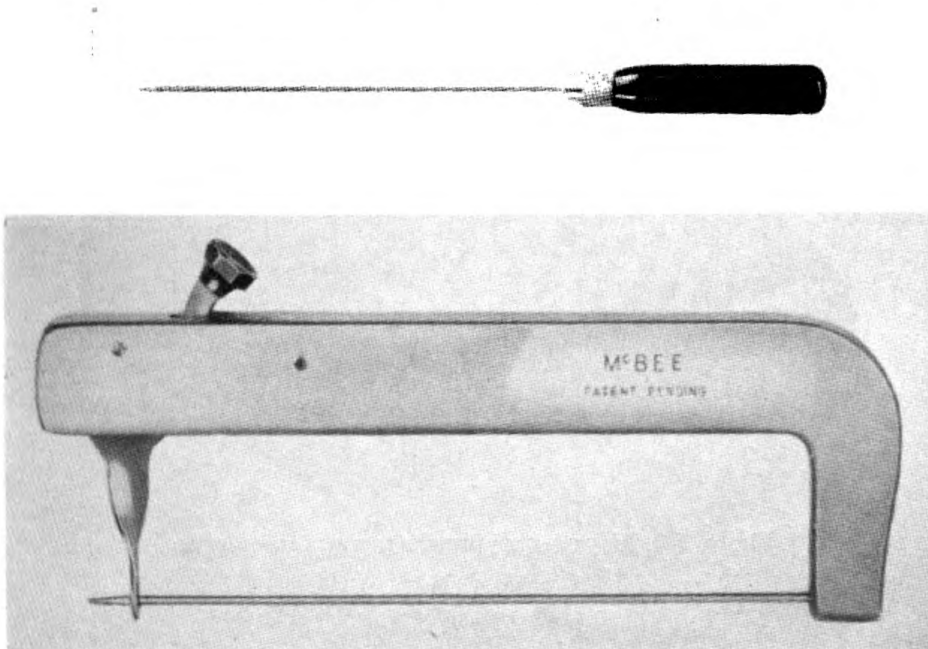


Figure 3-5. Keysorter, Types 5005 and 5006.

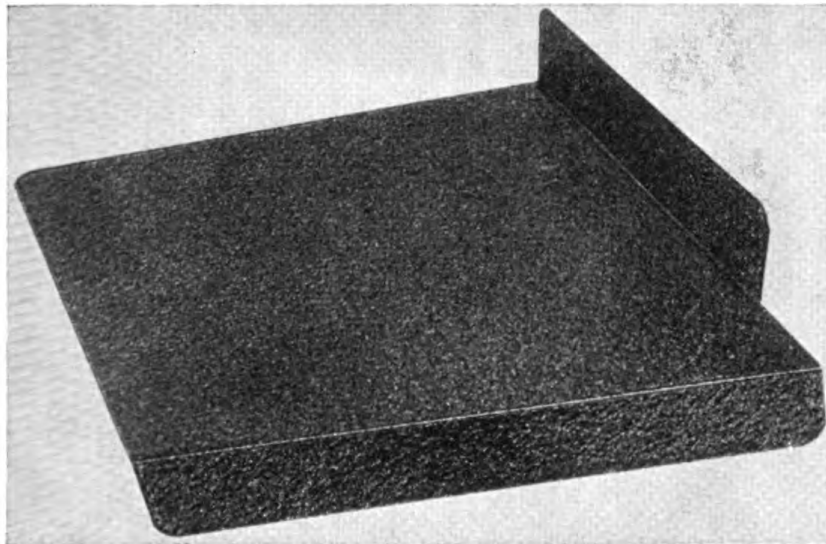


Figure 3-6. Keysort alignment block.

the block in the correct sorting position. The vertical guide along the right side of the block forms a right angle with the front edge of the desk. A rubber pad cemented to the bottom of the horizontal surface and the inside surface of the drop front prevents slippage during use.

The key punch (Figure 3-7) is used for punching Keysort cards one at a time, one entire edge being punched at each trip of the operating lever.



Figure 3-7. Keysort keypunch, manual or electric.

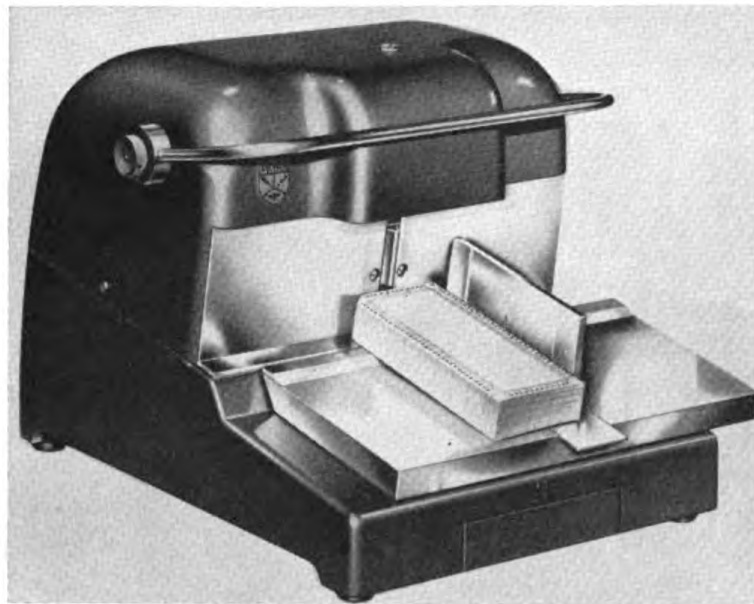


Figure 3-8. Keysort electric groover.

This machine, available either in manually or electrically operated styles, is limited to outer edge single-row punching, and is primarily designed for numerical coding in the 7-4-2-1 fields.

Grooving machines (Figures 3-8 and 3-9). As many as 100 cards, depending on the thickness of the stock, may be punched simultaneously in one position. This permits a considerable savings in time otherwise required for hand punching.

Keysort Selector (Figure 3-10). This device, which handles cards up to 10½ inches in length, makes a selective sort along an entire edge of the



Figure 3-9. Keysort manual card groover.

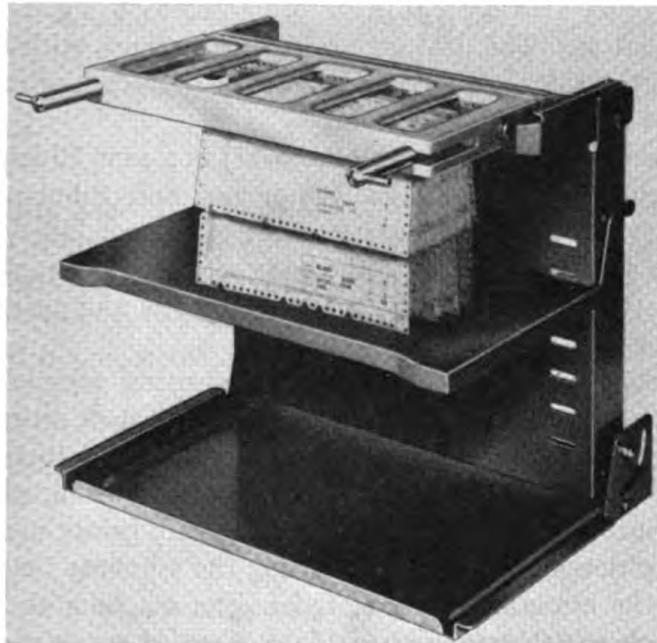


Figure 3-10. Keysort Selector.

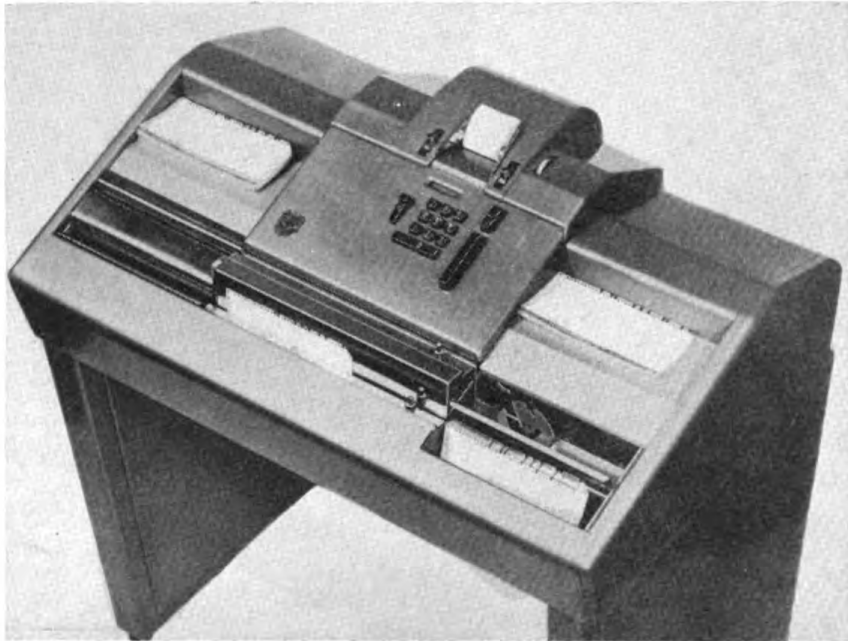


Figure 3-11. Keysort Tabulating Punch.

cards. About 200 cards can be handled at one time. A few simple selector codes suitable for use with this machine are described in Chapter 2. A reasonable speed is 60,000 multiple hole sorts per hour. When this selector is not in use, it may be folded and stored in a desk drawer.

McBee Keysort storage cabinets are specially designed for use with Keysort installations. The drawers are light in weight, with handles front and rear, and the sides are cut down so that they can be used as utility trays when the system is in operation.

Keysort tabulating punch (Figure 3-11) is a new 10-key adding and printing punch which automatically punches and tabulates quantities recorded in Keysort cards. It does a number of things:

- (1) Punches two quantities into the body of a Keysort card while simultaneously printing such quantities on a detail tape and accumulating the amounts for totaling.
- (2) Automatically senses (reads) the quantities, punched as above, from the Keysort card and simultaneously prints such sensed quantities on a detail tape, thereby accumulating the amounts for totaling.
- (3) Automatically reproduce-punches quantities from a pre-punched Keysort card into a blank Keysort card, and again prints such reproduced quantities on a detailed tape, accumulating the amounts for totaling.
- (4) Totals the amounts accumulated in the machine and, as desired, summary punches such totals into a blank Keysort summary card. Prints such totals on the detail tape.
- (5) Can be used as an ordinary 10-key adding machine if desired.



Figure 3-12. Keysort Data Punch.

Keysort data punch (Figure 3-12) simultaneously prints data on a punched card and marginally notches the coded information into the Keysort card. Metal plates of the Addressograph sort, embossed and notched with an employee's name and code, for example, are inserted into the data punch and with one stroke of the lever the data are both printed and punched.

The E-Z Sort System

E-Z SORT SYSTEMS, LTD.

45 Second St., San Francisco, Calif.

In this system, all coding and punching is done in a half-inch strip around the perimeter of the card. Cards are available with a single row, double row, triple row or quadruple row of holes, or combinations thereof. The single-row hole card employs an oval-shaped hole. By staggering the

row, it is possible to place six holes per inch and eliminate any projection when adjacent holes are grooved. The multiple-row card effects its saving in space by using smaller holes and projections between adjacent coded holes are eliminated by the special pattern cut out by the punch or groover. The entire system may be handled manually, although electric punches and groovers are available. Four to six hours instruction, combined with a few days practice, will generally produce an efficient operator. Any E-Z Sort machine can be mastered within a couple of hours. All the equipment is portable and designed for simple operation.

The system is intended for recording specific information on the card. Its expansion is unlimited, provided filing space is available. Present installations vary from those using 10,000 cards per day to those using 10,000 per year. The cards are durable; one installation has been in operation for over fifteen years. All types of E-Z Sort cards can be mended with a patented glued strip which is claimed to make the repaired card stronger than before mending.

E-Z Sort cards (Figure 3-13) are available in standard sizes ranging from $1\frac{3}{4} \times 4$ inches, in perforated strips containing four to sixteen cards, to sizes up to and including $8 \times 10\frac{1}{2}$ inches, containing a total of 205 holes on single-row hole cards. Cards larger than $8 \times 10\frac{1}{2}$ inches can be furnished by special processing. The number of cards that can be filed per inch varies from 80 to 150, depending upon the stock. A complete variety of card stocks can be obtained, including bond paper for sales tickets, etc., that are sorted once or twice and then filed or discarded, as well as special tag stocks and heavy durable rag content bristols for permanent records

The diagram illustrates the layout of an E-Z Sort card, which is a rectangular card with a grid of holes. The card is divided into several sections for data entry and indexing:

- Top Section:** Contains fields for **CENTURY**, **DECADE**, and **YEAR**, each with a row of holes for data entry.
- Right Section:** Labeled **AUTHOR & SOURCE INDEX**, featuring a grid of holes for alphabetical indexing.
- Bottom Section:** Labeled **LITERATURE CLASSIFICATION & RESEARCH INDEX**, also featuring a grid of holes for alphabetical indexing.
- Left and Right Margins:** Both margins contain vertical columns of holes, likely for additional sorting or filing purposes.

The card is designed to be used with a sorting machine, where the holes are punched into the card to categorize and sort documents.

Figure 3-13. E-Z Sort card.

and frequent sortings. E-Z Sort systems cards are also manufactured of special-purpose stocks such as Ozalid and other opalescent or sensitized materials.

E-Z Sort's patented hole structure, with depths up to four rows, is suited to direct-word coding with superimposed entries and is used extensively in research projects and technical literature files. Another feature is the ability to direct-sort letters or numerals up to four digits using a single sorter for each digit without incurring false drops. On three or four row holes, this particular coding arrangement uses fewer columns of holes than any other direct extraction arrangement.

Other multi-row hole arrangements allow the coding of the largest number of non-exclusive items in the least amount of perimeter space.

A variety of standard stock cards are available for the small installation including five sizes of analysis cards with one or two rows of holes, five varieties of bibliographic index cards with two, three or four rows of holes, and several miscellaneous cards, such as forms control, mailing list control, and employment cards.

Manual equipment (Figure 3-14) needed for the simplest type of operation consists of the hand groover, sorting needles and sorting tray which is used for holding cards while sorting. The hand groover permits the operator to code all four sides of the card at the rate of about 180 per hour. The hand groover is manufactured in several styles, including an intermediate type for cutting between the rows of holes on multiple row hole cards. The groover blade is replaceable when dull, and has a patented locator tip to assure proper register to the desired depth when coding. For example, the Model P-4 hand groover will groove one, two, three or four rows deep by inserting the locator tip into the hole being grooved. All of this equipment is offered for sale.

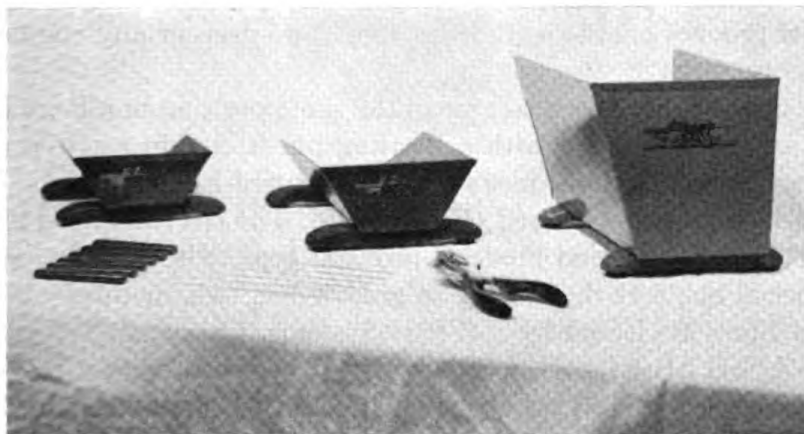


Figure 3-14. E-Z Sort hand groover, sorting needles, and sorting tray.

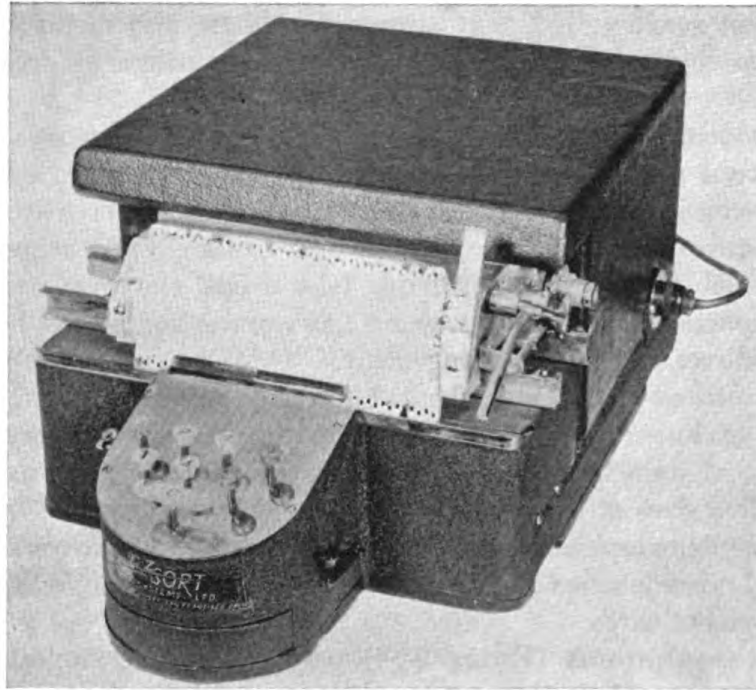


Figure 3-15. E-Z Sort electric key-groover.

The **E-Z Sort electric key-groovers** are available in two basic models: Model 1 (Figure 3-15) is a tabulating carriage type designed to groove 7, 4, 2, 1, 0 or other codings on the single-row hole card. With only seven keys all four edges of 500 to 600 cards per hour can be grooved.

Model 2 (Figure 3-16) is designed to groove one entire edge of the card in a single operation. Keys can be set to repeat any part or all of the coding on succeeding cards without resetting the keys each time. This key-groover is manufactured in three styles—one row, two row or four row grooving depths. The two-row model grooves one or two rows deep and the four-row model grooves one, two, three or four rows deep or any combination, as may be required.

Batch groovers are used to groove the same position in a large number of cards and are equipped with a locating pin to assure positive registry when grooving. These machines are manufactured in two models: electric (Figure 3-17) and foot-powered (Figure 3-18). The electric model is a desk type and the foot-powered model a pedestal type. They can be equipped with a special blade at the factory to groove one, two, or three rows deep. These machines are leased by E-Z Sort Systems, Ltd.

The **E-Z Sort Multi-sorter** (Figure 3-19) is a selector unit. The needles are positioned according to the particular selective sort required. The unit may be set up in approximately one minute to separate cards coded in the required pattern along a single edge of the cards. It can handle 100 to 300



Figure 3-16. E-Z Sort electric key-groover, for whole edge of card.



Figure 3-17. E-Z Sort electric batch groover.

cards at a time for a total of more than 36,000 cards per hour. Figure 3-19 also illustrates the use of one of the catching trays. This device is available to handle any of the E-Z Sort Systems multiple-row hole cards up to 11 inches long.

E-Z Sort Card Counter. This machine is small enough to be placed on a desk and will accurately count approximately 100,000 3 x 5 inch cards per hour. It can be leased.

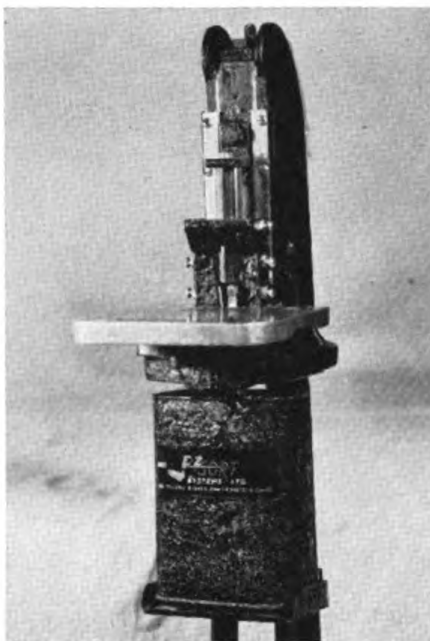


Figure 3-18. E-Z Sort foot-powered batch groover.

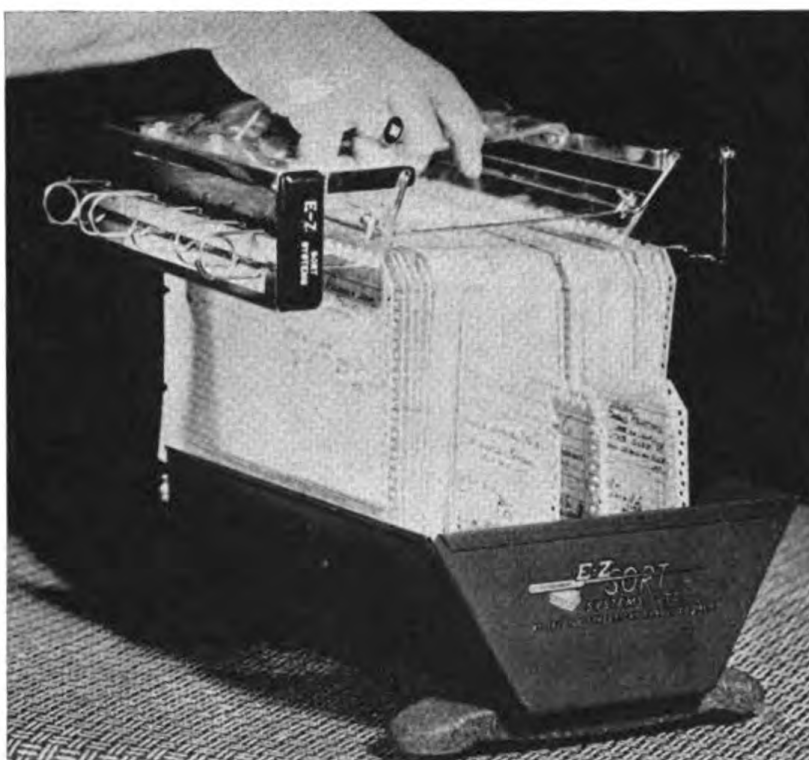


Figure 3-19. E-Z Sort Multi-Sorter and sorting tray.

The Unisort System

BURROUGHS CORPORATION, THE TODD COMPANY DIVISION
Rochester 3, N. Y.

CHARLES R. HADLEY PLANT, *Los Angeles 12, Calif.*

UniSort is the registered trademark for The Todd Company line of edge-punched cards. UniSort cards are marketed through Todd-Hadley representatives in all parts of the United States, Canada, Alaska, and Hawaii.

UniSort systems are manually operated and all phases of operation, including operation of the equipment, can be learned in a very short time. Maximum operating speed on the keyboard notching machine and maximum sorting speed with the cards can be attained within a few days.

All notching equipment, with the exception of hand notchers, is available on a lease basis only.

Standard **UniSort cards** (Figure 3-20) are available in sizes ranging from 3 x 5 inches to 6 $\frac{5}{8}$ x 8 $\frac{1}{2}$ inches; special size cards can be made to meet the requirements of the system in which they are to be used. There is no standard UniSort card for bibliographic use. The card shown is a special one printed for a research institute. It illustrates triangular coding, so often used because it facilitates selective sorting. A deep hole notch on the left side and a shallow hole notch on the right, where the lines intersect, indicate the notching for the top letter or number in the square. Reversing the deep and shallow notches indicates the notching for the bottom letter or number.

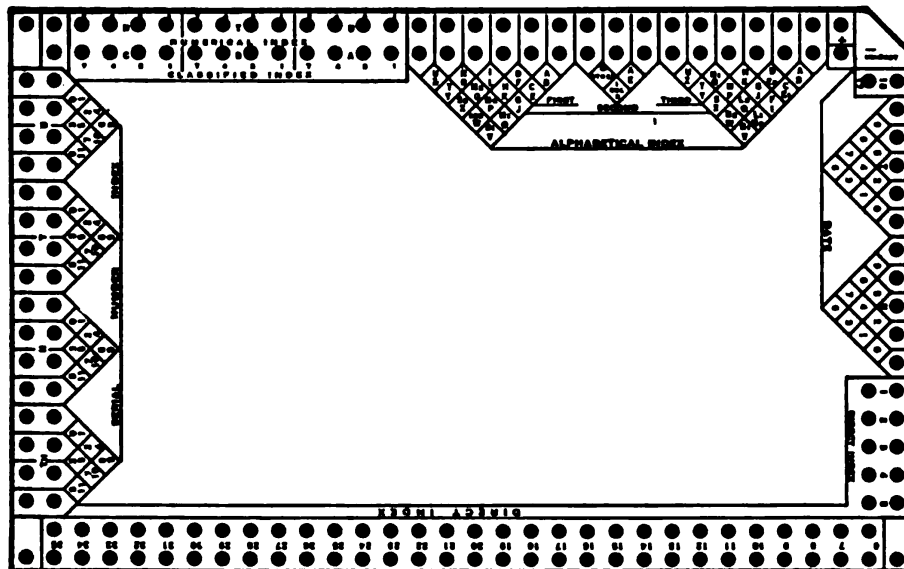


Figure 3-20. UniSort card.



Figure 3-21. UniSort keyboard notching machine.

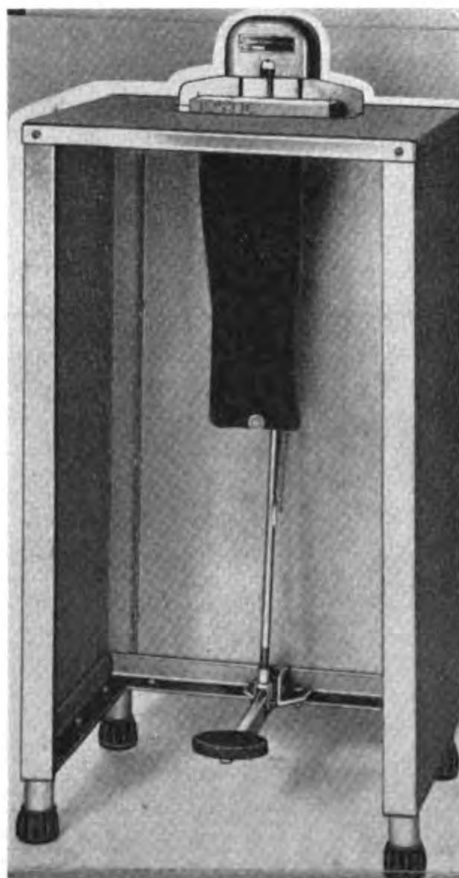


Figure 3-22. UniSort foot-powered notching machine.

The **Keyboard notching machine** (Figure 3-21) is used for notching up to three UniSort cards at a time, an entire edge being notched with one depression of the motor bar. Cards are fed into the front of the machine. This machine is available either in hand or electrically operated styles, for either four or five-hole-to-the-inch cards.

The **Foot-power notching machine** (Figure 3-22) will notch the same hole in 200 cards at a time. Identical information (such as the date) may thus be notched into the edge of thousands of cards in a very short time.

The **desk model gang notcher** (not illustrated) will notch the same hole in 50 cards at one time. This machine is used where the volume of cards is too small to warrant the use of a foot-power notching machine.

UniSort alignment discs (Figure 3-23a) are used to hold a stack of cards in alignment while notching them on the foot-power notching machine or the desk model gang notcher, thus attaining a more perfect notch for all cards in the stack.

UniSort hand punches are available in four different styles: (1) ticket style (Figure 3-23b) for punching in the body of the card; (2) deep hole with gauge (Figure 3-23c); (3) shallow hole, with gauge and receptacle (Figure 3-23d); and (4) shallow hole, with gauge (Figure 3-23e).

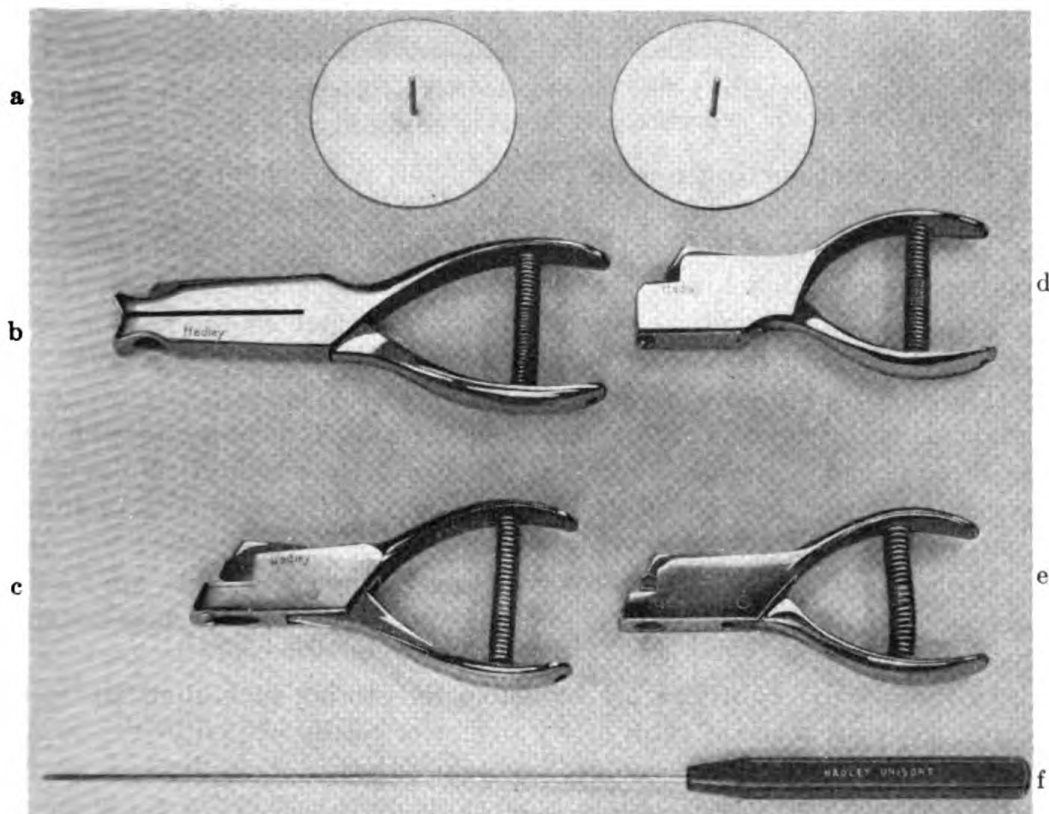


Figure 3-23. UniSort alignment discs, hand punches, and sorting needle.



Figure 3-24. UniSort card holder, shown on back of keyboard notching machine.

The **UniSort sorting needle** (Figure 3-23f) is a lightweight, stainless steel needle set in a molded plastic handle.

The **UniSort card holder** (Figure 3-24) slides into brackets on the back of the keyboard notching machine. It holds a stack of cards in such a position that the operator can read the top card and set the keys before removing the card to insert it for notching.

The **Sorting Pan** (Figure 3-25) speeds the sorting operation. The raised guide along the right side of the pan stops the cards that have dropped out as those remaining on the needle are lifted over the guide. A rubber pad, cemented to the bottom of the pan, prevents it from slipping or from marring the desk.

UniSort pull tubs, used as reservoirs for pre-punched cards, are available in two capacities, 9,500 or 15,000 cards. These are of steel construction mounted on rubber-tired casters.

UniSort card cabinets are available for storing and filing UniSort cards. They are made of extra heavy steel, electrically welded, with a gray baked enamel finish. The cabinets contain either 8 or 16 drawers, each with a capacity of approximately 1,250 cards.

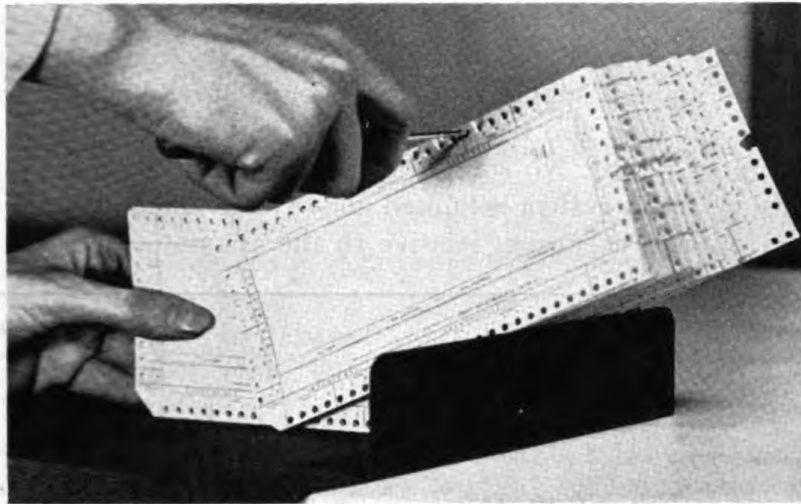


Figure 3-25. UniSort sorting pan.

The Findex System

WILLIAM K. WALTHERS, INC.

1245 N. Water St., Milwaukee 2, Wis.

The Findex System is provided with cards having perforations equally spaced in the body and/or the edge. Space is provided at the top and on the reverse side of the cards for written records. Information is coded by cutting slots between adjacent vertical perforations. These slots are used in separating the cards by means of steel rods.

Operation is entirely manual. The cards are typed, marked for slotting, slotted by hand, and filed vertically in any manner convenient to the user—for instance, alphabetically. They may be kept in an ordinary drawer or in the selector which is used when the cards are to be separated. The selector is a steel drawer, the front and rear plates of which contain perforations matching those of the cards. Steel rods are inserted through the proper perforations, as determined by the method of coding, and extend the full length of the drawer. When the drawer is inverted and the cards stroked or shaken, cards in which there are slots corresponding to the rods drop a distance equal to the length of the slot. All cards are then locked in place by means of rods located at locking positions and the drawer is re-inverted. The selected cards may then be inspected or removed as desired. Each selector will accommodate 600 cards. It requires about three minutes to set the unit up for segregation; therefore approximately 12,000 cards may be sorted per hour. If the cards are not removed, no refiling is necessary. If they are removed, the refiling may be expedited by placing a colored slip in the file where each card has been removed.

PUNCHED CARDS

No particular skill or training is required. Anyone who is capable of typing the cards can handle the sorting and other operations after reading the operator's manual supplied by the manufacturer. All equipment is sold outright but may also be obtained on a rental basis, subject to sale within a limited period. No special maintenance problems are involved. The expansion of this system is limited only by the amount of filing space available. However, expansion relative to the amount of information to

NAME _____ ADDRESS _____ MAN TO ADDRESS: _____		R.S.S. Salesman																																																	
DO NOT ADDRESS BELOW THIS LINE																																																			
PRODUCTS																																																			
ADVERTISING AGENCY																																																			
REMARKS:																																																			
		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 5%;">A</td> <td style="width: 5%;">P</td> <td></td> </tr> <tr><td></td><td></td><td>RAILWAY AGE</td></tr> <tr><td></td><td></td><td>MOTOR TRANS.</td></tr> <tr><td></td><td></td><td>MARCH DAILY</td></tr> <tr><td></td><td></td><td>JUNE DAILY</td></tr> <tr><td></td><td></td><td>MECH. ENG.</td></tr> <tr><td></td><td></td><td>ELEC. ENG.</td></tr> <tr><td></td><td></td><td>SIGNALING</td></tr> <tr><td></td><td></td><td>MAINTENANCE</td></tr> <tr><td></td><td></td><td>CAR BLD. CYCL.</td></tr> <tr><td></td><td></td><td>LOCO. CYCL.</td></tr> <tr><td></td><td></td><td>MAIN. CYCL.</td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>		A	P				RAILWAY AGE			MOTOR TRANS.			MARCH DAILY			JUNE DAILY			MECH. ENG.			ELEC. ENG.			SIGNALING			MAINTENANCE			CAR BLD. CYCL.			LOCO. CYCL.			MAIN. CYCL.												
A	P																																																		
		RAILWAY AGE																																																	
		MOTOR TRANS.																																																	
		MARCH DAILY																																																	
		JUNE DAILY																																																	
		MECH. ENG.																																																	
		ELEC. ENG.																																																	
		SIGNALING																																																	
		MAINTENANCE																																																	
		CAR BLD. CYCL.																																																	
		LOCO. CYCL.																																																	
		MAIN. CYCL.																																																	
PROMSPECT Shop Equip. 1 ● 11 ● 21 ● 31 ● Motor Transport 2 ● 12 ● 22 ● 32 ● base Daily 13 ● 23 ● 33 ● Mechanical Engineer 4 ● 14 ● 24 ● 34 ● Electrical Engineer 5 ● 15 ● 25 ● 35 ● Railway Signaling 6 ● 16 ● 26 ● 36 ● Eng. & Main-tenance 7 ● 17 ● 27 ● 37 ● Car-Bld. Cyclopedic 8 ● 18 ● 28 ● 38 ● Loco-motive Cyclopedic 9 ● 19 ● 29 ● 39 ● Main-tenance Cyclopedic 10 ● 20 ● 30 ● 40 ●		ARTER-TRAILER Railway Age 31 ● March Daily 32 ● base Daily 33 ● Mechanical Engineer 34 ● Electrical Engineer 35 ● Railway Signaling 36 ● Eng. & Main-tenance 37 ● Car-Bld. Cyclopedic 38 ● Loco-motive Cyclopedic 39 ● Main-tenance Cyclopedic 40 ●																																																	
SIMMONS-BOARDMAN PUBL. CORP. New York																																																			
Index. Milwaukee																																																			

Figure 3-26. Findex card.

be coded or recorded on the cards is limited by the size of the card. **Cards** (Figure 3-26) are available in two sizes—6 x 8 and 8 x 8 inches. They are made of 0.012-inch ledger stock with a high linen content and are durable enough to outlast ordinary usage. Wrong coding may be corrected by pasting a small linen "patch" over the slotted portion between the two holes. The intended use of the card will determine how much of its area is to be used for punching and how much is to be used for written records. No provisions have been made at present for microfilm insertions or photosensitization.

Selectors. The selector drawers, which may also act as permanent filing space, house 600 cards each. They may be handled singly on a revolving "cradle" or in cabinets of two, four, six, or eight selectors each. These cabinets are provided with special slides which permit inversion of the drawers during the selecting process without removing them from the cabinet. A cabinet holding eight such drawers (4,800 cards) occupies three to four square feet of floor space, depending on the size of the card.

The only other implements required are the steel sorting rods and the slot punch which resembles a desk-type paper punch. All of the equipment is portable. The system appears relatively inexpensive and well adapted to the requirement for a large number of cards and a comparatively small amount of information on each card. It is primarily designed for records which require group analysis, correlation, and cross indexing, when it is desirable to have all information on one card. It does not lend itself to tabulation of accounting facts or to serial sorting.

The Flexisort System

SUPERIOR BUSINESS MACHINES, INC.

285 Madison Ave., New York, N. Y.

The Flexisort system operates upon the same basic principles as the marginal hole punched card systems described previously but it is unique in its use of a punch which automatically cuts all necessary holes into unperforated cards. Such a machine eliminates the need for specially prepared cards. Any kind, size, or weight card stock can be used. It also permits conversion of existing index card files into marginal hole punched-card files.

Flexisort cards (Figure 3-27) may be prepared either singly, as one of the parts of a set of precollated snap-out forms, or they may be prepared in continuous form. When prepared in continuous form, it is possible to type an invoice or listing as the original copy and simultaneously produce, as a carbon copy, a separate card for each entry on the list.

Flexisort punch (Figure 3-28). This machine is electrically operated and resembles a desk-type adding machine with an eight-bank keyboard.

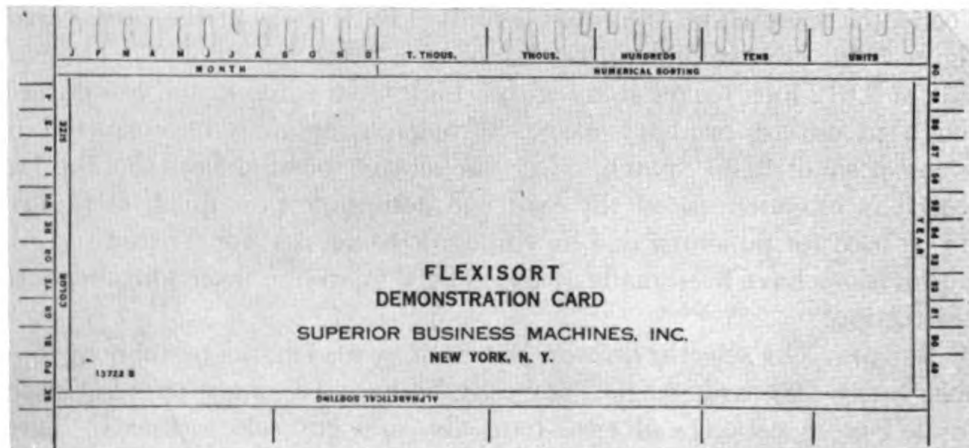


Figure 3-27. Flexisort card.



Figure 3-28. Flexisort electric punch.

Each bank contains a correction key, numerical keys numbered 1 through 9, a repeat key, and four supplementary keys which permit the direct coding of miscellaneous statistical information. Every second column of keys carries alphabetical characters as well as numerals. The cards are

perforated and punched by depressing the keys corresponding to the positions to be coded and then pressing down the motor bar. The single stroke perforates the holes and punches the notches along one edge of the card. All eight columns may be set prior to the simultaneous cutting operation. When a numerical key is pushed, four dies are arranged within the machine and when activated they punch a pattern into the margin of the card corresponding to the 7-4-2-1 code previously described. Each number from 1 to 9, inclusive, can be punched with a single key depression, but for alphabetical punching two keys are depressed for each letter. For example, to code "B," the key marked with the letters "ABC" and the key "2" (signifying the second letter of that group) in the column immediately to the right are both depressed. Each of the keys provided for direct coding controls only one die and may not be used simultaneously with the alphabetical or numerical keys in the same column. Erroneously set keys may be released by depressing the correction key. The repeat keys lock the dies into position and make it possible to cut common information into successive cards without resetting the board. As many as three cards can be gang punched simultaneously. Provision is made for suppressing punching in areas where it is not desired.

The Needlesort System

ARIZONA TOOL & DIE COMPANY

31 E. Rillito St., Tucson, Ariz.

Needlesort cards (Figure 3-29) are manufactured in two sizes, 3½ x 6 inches and 5 x 8 inches. They are only punched along three edges—the

2 1 7 4 2 1		1 7 4 2 1		7 4 2 1		7 4 2 1		1 7 4 2 1		7 4 2 1		R 1			
C. TEST STAGE				D. SYSTEM				E. COMPONENT				F. ELEMENT			
SHOP SERIAL:				MISSILE/EQUIPMENT TYPE:				FAILED IN OR DUE TO:				☐ STORAGE ☐ HANDLING ☐ INSPECTION ☐ ATTEMPTED LAUNCH ☐ TEST/CHECK ☐ FLIGHT			
DATE:				MISSILE/EQUIPMENT SERIAL:				ESTIMATED ACCUMULATED OPERATING TIME OF COMPONENT AT FAILURE:				PLATES "ON" MINS. FILAMENTS "ON" HRS. ENERGIZED HRS.			
WHAT WERE THE SYMPTOMS OF FAILURE?															
ITEM:		NAME, MODEL, OR TYPE:		SERIAL, PART NO., OR SYMBOL		MEASURED OUTPUT OR CONDITION BEFORE REPAIR		AFTER REPAIR							
COMPONENT OR UNIT															
SUB-ASSEMBLY OR CIRCUIT															
INDIVIDUAL PART OR ELEMENT															
WHAT WAS WRONG WITH THE PART?															
WHAT DID YOU DO TO REPAIR THE COMPONENT?															
PROBABLE CAUSE OF FAILURE (REMARKS OR RECOMMENDATIONS)															
MAN-HOURS REQUIRED TO REPAIR		☐ SEE OTHER SIDE FOR ADDITIONAL DATA OR DIAGRAMS		CHECKED AND APPROVED BY:		DATE:									
DEFECT CARD:															

Figure 3-29. Needlesort card.

two sides and the top. The 5 x 8 inch cards are available in three colors, buff, pink or yellow, and can be imprinted or mimeographed by the company. A notching punch and sorting needle are also offered for sale.

The Zatocoding System

ZATOR COMPANY

140½ Mt. Auburn St., Cambridge 38, Mass.

The Zatocoding System consists of the Zator Selector, edge-notched Zato cards, the techniques of using random descriptor code patterns notched in superimposition along the edge of the cards, and the use of descriptors by which documents are characterized. Deriving descriptors is an empirical process. Chapter 15 describes this process for a given installation.

The **Zator 800 Selector** (Figure 3-30) holds an easy handful of Zato cards. The box is vibrated by a small motor. Rods or needles running from front to back of the box are inserted in a desired selective pattern. Cards notched in the positions corresponding to the selector rods fall down from the rest of the pack and can be easily separated. With the 800 Selector, sorting speeds of about 800 cards per minute can be attained.

Zato cards come in two styles, one with notches along a single edge and the other with notches along two edges (Figure 3-31). The latter type has 72 notching sites as compared to 40 on the single-edge style.

For the Zatocoding system, a list of random-like patterns has been prepared. Any patterns that are random-like, in the sense that the individual code marks are well scattered and fall with approximately equal incidence on all the coding sites, can be used for coding.

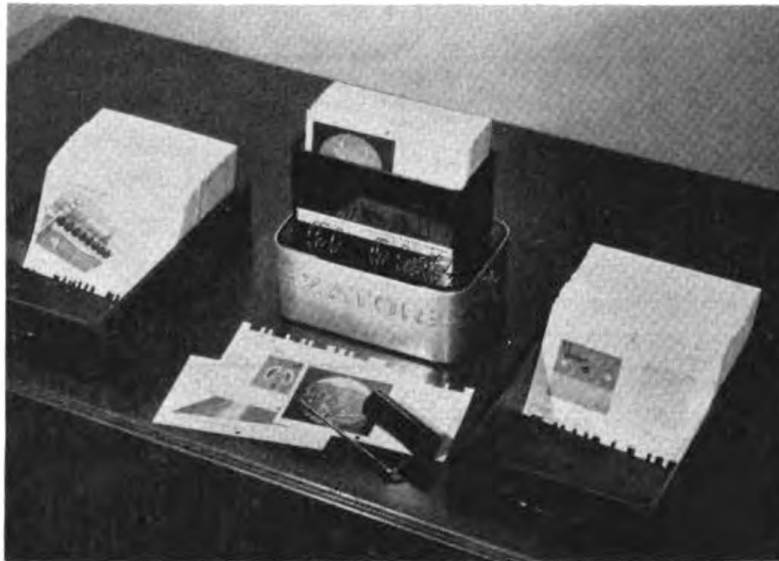


Figure 3-30. Zator 800 Selector.

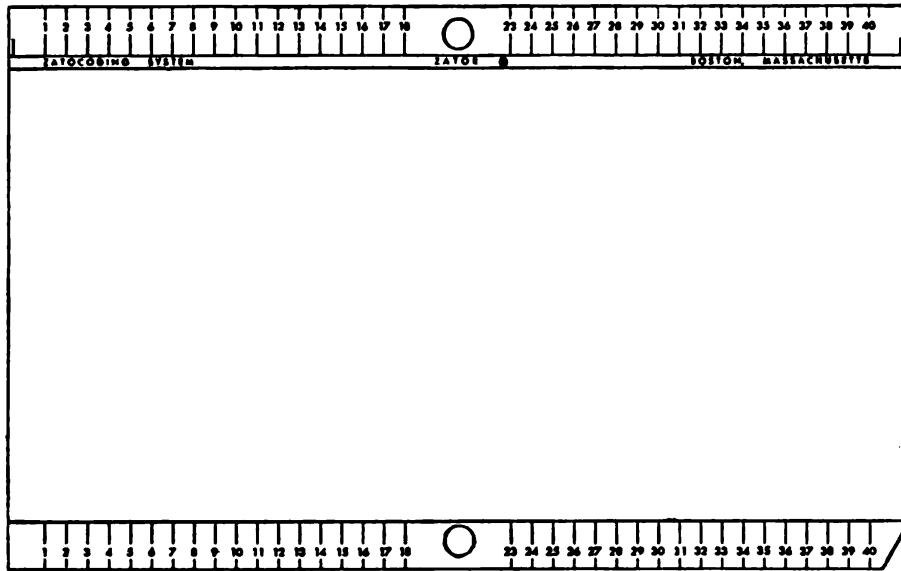


Figure 3-31. Zatocard.

Technical consultation and guidance are provided by the Zator Company during the installation of a commercial Zatocoding system. The system is provided on a rental basis, including a license for making the randomly coded cards.

Foreign Manufacturers

Edge-notched punched cards are manufactured by a large number of firms outside the United States. The list that follows is by no means exhaustive.

Buro-Organisation G. m. b. H.
Brandenburgische Str. 27
Berlin W 15, Germany

Esselte System
Tandlarkarhogskolan
Malmo, Sweden

Copeland-Chatterson Company,
Ltd.
Exchange House
Old Change
London E. C. 4, England

National Luchtvaartlaboratorium
Amsterdam
Sloterweg 145
Amsterdam-W., Netherlands

Edler & Krische
Kestner Str. 42
Hannover, Germany

Rapidtri
78, Rue de Wattignies
Paris 12^e, France

Eichhoff-Werke G. m. b. H.
Dieffenbach Str. 2
Schlitz/Hessen, Germany

G. Schmid Verlag
Herderstrasse 2
Lubeck, Germany

TABULATING TYPE PUNCHED CARDS

The IBM System

INTERNATIONAL BUSINESS MACHINES CORPORATION

590 Madison Ave., New York 22, N. Y.

IBM accounting machines are widely used not only in commercial accounting and statistics but also for processing many kinds of scientific data. In laboratories throughout the country, these machines are used to create, maintain, and search punched card reference files and to perform mathematical computations necessary in solving many problems in astronomy, ballistics, chemistry, engineering, meteorology, and physics.

Once the initial data have been recorded as holes in the cards, the machines can sense these holes electrically and automatically perform a wide variety of operations such as rearranging the cards into any required sequence, transferring data from one card to another, printing the information on the cards or on a sheet of paper, consulting tables of data, and performing the arithmetical operations of addition, subtraction, multiplications, and division. Electrical impulses, transmitted through the holes in the cards, are used to read the recorded data and control the operation of the machines.

Branch offices, service bureaus, and representatives are located in principal cities. Equipment can be bought or supplied on a monthly rental basis which includes the use and maintenance of the machine. For those who do not have the necessary installations, the Service Bureau Corporation, a wholly owned subsidiary corporation, offers services for the preparation of reports on a time or complete job basis.

The company maintains a training program for the representatives of its customers. Classes are held at IBM offices and educational centers for instructing the customer personnel in key punching, operation of the various machines, control panel wiring, and other related subjects. Selection of the personnel to receive instruction is determined by the customer. Classes also are held for supervisors and managers of machine accounting departments. At Endicott and Poughkeepsie, New York, IBM conducts classes for customer executive personnel.

Cards (Figure 3-32) are supplied in a variety of sizes but the most widely used one is $3\frac{1}{4} \times 7\frac{3}{8}$ inches. A stack of approximately 150 cards measure one inch in thickness. The $3\frac{1}{4} \times 7\frac{3}{8}$ inch card contains 80 vertical columns divided into 12 punching positions. Combinations of these 12 punching positions are used to punch alphabetical and numerical information in the card.

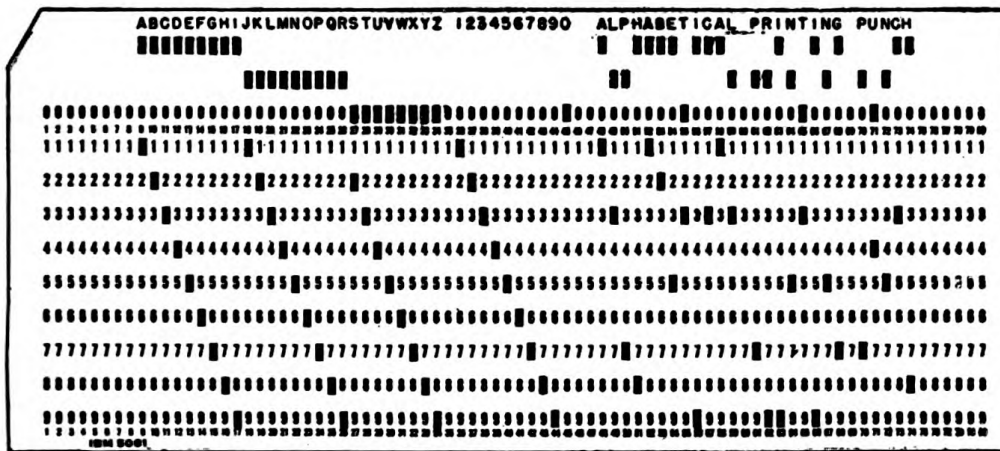


Figure 3-32. IBM card.



Figure 3-33. IBM card punch, Type 24.

Key Punches are used for punching the holes in the cards. A number of different machines can perform this function. The card punch (Figure 3-33) is used for recording both alphabetical and numerical information, and is equipped with a combination keyboard designed for high-speed operation. Cards are fed into the machine automatically and move forward column by column under the control of a program card which governs duplicating, skipping, and the kind of information (either alphabetical or numerical) to be punched into the cards from the combination keyboard. A duplicating

1 206 W V ASTUR										J 375 1231										JOB TIME CARD																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
DEPT. EMPLOYEE NO.					NAME OF EMPLOYEE					HOURLY RATE					DATE																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
NAME OF EMPLOYEE										ACCOUNT OR JOB NO.					PART NO.					OPER. NO.					MACH. NO.					PIECES FINISHED					HOURS					KIND																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
☐ - NO MASTER ☐ - ABSENT ☐ - NEW EMPLOYEE ☐ - SUSPENDED										01					02					03					04					05					06					07					08					09					10					11					12					13					14					15					16					17					18					19					20					21					22					23					24					25					26					27					28					29					30					31					32					33					34					35					36					37					38					39					40					41					42					43					44					45					46					47					48					49					50					51					52					53					54					55					56					57					58					59					60					61					62					63					64					65					66					67					68					69					70					71					72					73					74					75					76					77					78					79					80					81					82					83					84					85					86					87					88					89					90					91					92					93					94					95					96					97					98					99					100					101					102					103					104					105					106					107					108					109					110					111					112					113					114					115					116					117					118					119					120					121					122					123					124					125					126					127					128					129					130					131					132					133					134					135					136					137					138					139					140					141					142					143					144					145					146					147					148					149					150					151					152					153					154					155					156					157					158					159					160					161					162					163					164					165					166					167					168					169					170					171					172					173					174					175					176					177					178					179					180					181					182					183					184					185					186					187					188					189					190					191					192					193					194					195					196					197					198					199					200					201					202					203					204					205					206					207					208					209					210					211					212					213					214					215					216					217					218					219					220					221					222					223					224					225					226					227					228					229					230					231					232					233					234					235					236					237					238					239					240					241					242					243					244					245					246					247					248					249					250					251					252					253					254					255					256					257					258					259					260					261					262					263					264					265					266					267					268					269					270					271					272					273					274					275					276					277					278					279					280					281					282					283					284					285					286					287					288					289					290					291					292					293					294					295					296					297					298					299					300					301					302					303					304					305					306					307					308					309					310					311					312					313					314					315					316					317					318					319					320					321					322					323					324				

Figure 3-34. IBM “Mark-sensed” card.

feature permits automatic punching of common information from one card into the next. A printing card punch is available that performs all the functions of the card punch and which also prints along the top edge of the card the characters that are coded in its columns.

Cards can also be punched automatically, without the use of key punching, through a medium known as “mark-sensing”. This embodies the use of a graphite pencil to make short marks directly on the face of the card. The cards marked in this manner (Figure 3-34) are fed into another machine which electrically senses the graphite marks and punches corresponding holes into the desired position on the same card.

Still another way to enter original data into cards is by means of the **Typewriter Card Punch**, which will simultaneously type a document and punch cards with selected data from the typewritten record. IBM also manufactures a typewriter tape punch which records selected data into a paper tape. This tape can be transmitted by mail or teletype to a remote location where a tape-to-card machine will convert paper tape to punched cards.

Verifiers detect transcription errors. Once the card is punched, it becomes the basic record from which all transcription is subsequently done by machine. However, since this does not relieve the possibility of errors in the original punching, the verifier has been provided to check punching accuracy. As in key punching, the verifier keys are depressed as if the same information were being recorded once more. If the punched positions in the card do not correspond to the keys depressed in the verifier, an error signal is made. It can then be determined whether the error is in the punching or in the verification. Cards are notched on the right hand edge as visible proof that they have been verified and are punched correctly; in

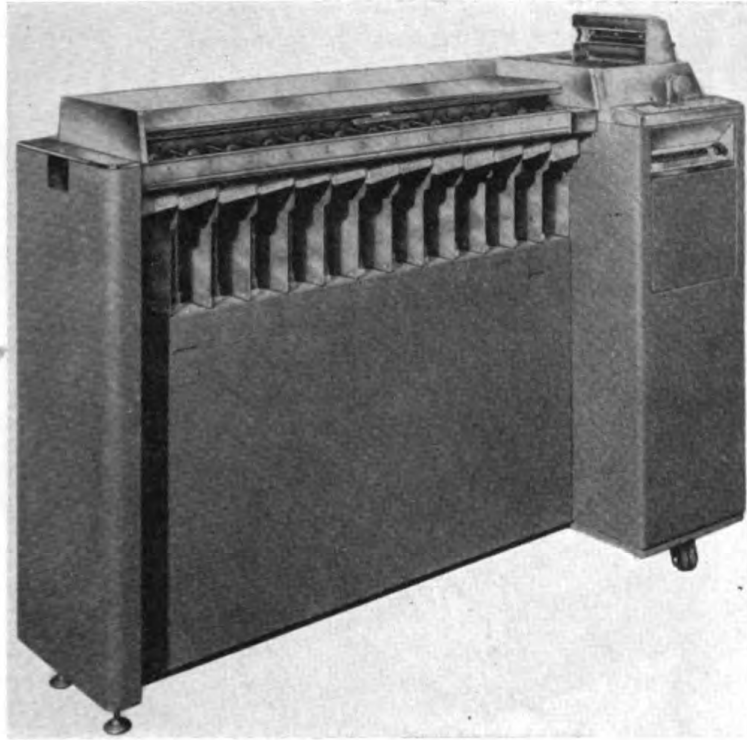


Figure 3-35. IBM sorter, Type 083.

the event of an error they are notched directly over the column in which the error appears.

A device known as the Self-checking Number Device used with the Key Punch may, in some cases, obviate the use of a verifier. With this device, the Key Punch will assign one extra digit to any code or identifying number. Thereafter, when that number is used, errors either in the original handwritten record or in the punching of the card are automatically revealed.

Sorters (Figure 3-35) arrange cards in any desired order according to the data punched into them. Also they separate the cards into groups having certain specific information. This sorting process takes place one column at a time, so that by successive sortings the cards may be arranged in any desired order. The 13 pockets into which cards can be distributed correspond to the 12 vertical punching positions in a card plus one pocket for cards having no holes in the column being sorted. If any given data are being selected, this last pocket receives those cards not having the desired information. The cards deposited into any one pocket remain in the same sequence in which they were fed into the machine. These units automatically stop when a pocket is filled. The one illustrated handles 39,000 sorts per hour.

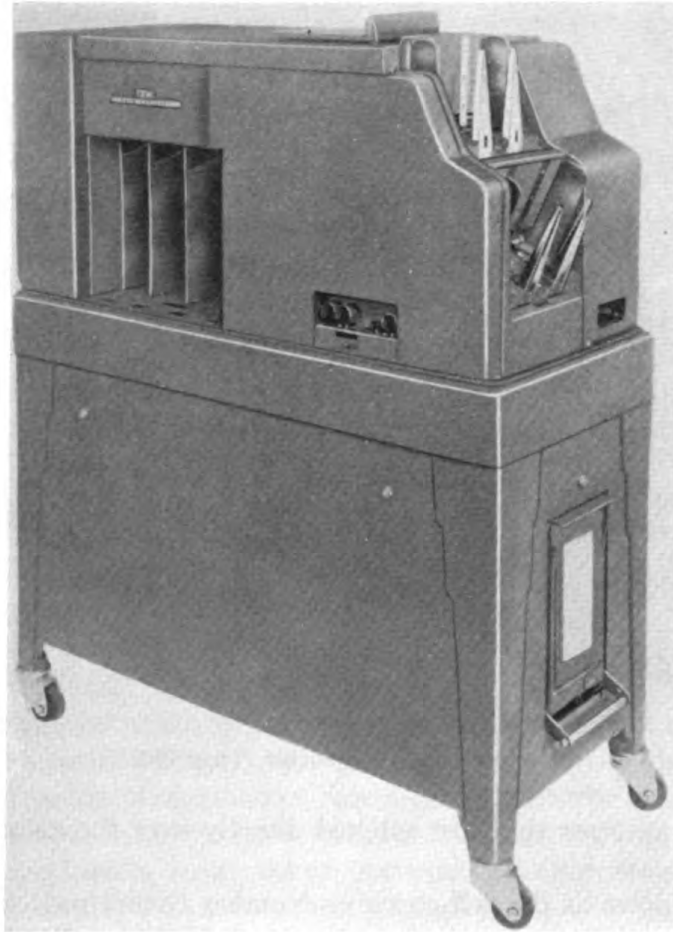


Figure 3-36. IBM collator, Type 077.

Collators (Figure 3-36). The principal function of the collator is to feed and to compare two sets of punched cards simultaneously, in order to match them or to merge them. While doing this the collator can separate the cards which match from those which do not, thereby making it possible to pull as well as to file cards automatically. There are two feeds, each of which operates at the rate of 14,400 cards per hour, and there are four pockets into which the cards can be separated, for example, into two groups of matched cards and two groups of unmatched cards. As a filing machine, the collator simultaneously feeds and compares two groups of cards. These two groups are merged into a single group in numerical or alphabetical sequence, or, if preferred, the cards may be matched into two identical groups. While either of these operations is being performed, the collator will remove from either group those cards that do not match, those which are out of sequence, or those which match cards in the other group or other selected cards.

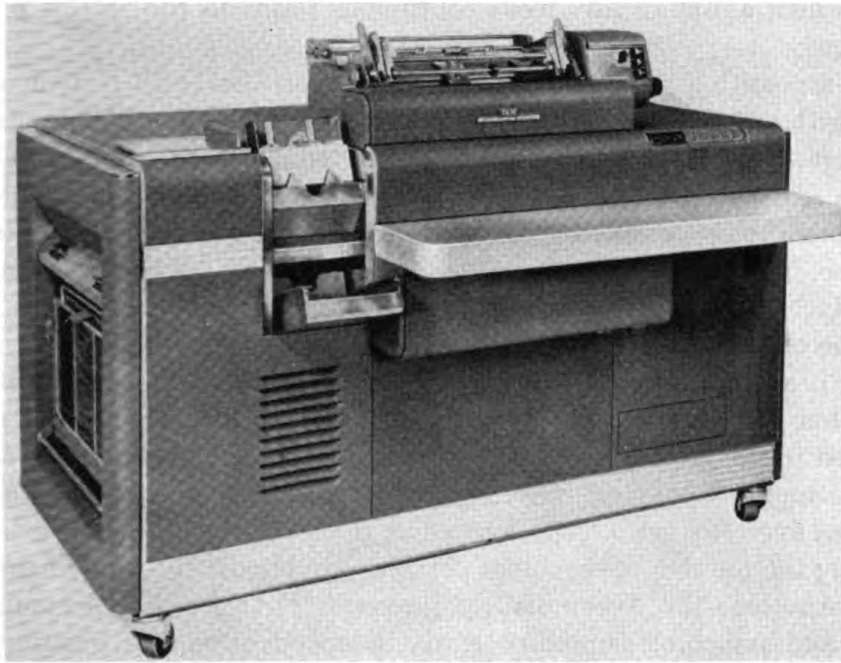


Figure 3-37. IBM alphabetical accounting machine, Type 403.

Printing Units. The IBM accounting machine is used to obtain printed reports of data punched in the cards. This is a machine which selects and reads data from cards, adds and subtracts, and prints on a sheet of paper data from individual cards or from accumulated totals. The machine will add or subtract as many as 112 digits at a time and print as many as 120 characters in a single line. There are several types of accounting machines, all basically the same in operation. Some of these print only numerical data, while others print both numerical and alphabetical data. Several models will print three lines from a single punched card.

Most widely used, however, is the machine (Figure 3-37) which prints one line of numerical or alphabetical information from each card. This unit will print 88 positions of numerical data, or 43 positions of alphabetical data on the left, and 45 positions of numerical data on the right. Every character on a line is printed simultaneously. The unit may be used to print selected details from every card or from specific cards. It also may be used to accumulate selected data and print the various classifications of totals. While several speeds are available, the fastest will perform detail printing at the rate of 9,000 lines per hour and will accumulate totals without detail printing at the same speed. The machine is equipped with major, intermediate, and minor classification controls which provide for the printing of group totals when any of these classifications change. Important to the high speed of this operation is the tape-controlled car-

riage, which automatically feeds continuous forms to the correct printing positions.

The accounting machine, like most others in the IBM system, is provided with a control panel on which pluggable connections can be made by the operator. In this way, for example, the number read from a given card column can be routed to any adding or printing unit. Other connections determine the operations to be performed as each card passes through the machine. The control panel can be easily removed for wiring changes or for replacement by another panel that has already been wired for a different type of operation.

Summary Punches. To record totals accumulated in the printing unit into other cards for use in subsequent operations, a punching unit can be attached to the accounting machine by means of a cable. There are a number of summary punches, several of which are also used as key punches; one goes even further and performs five different machine functions, thus reducing the number of machines necessary to handle relatively light work loads. Another, the Accumulating Reproducer, is used for accumulating totals and punching summary cards independently of a printing unit. This machine checks all of the accumulated totals as well as their punching.

Reproducers automatically transcribe punching from one card to another, thereby limiting clerical transcription to the original operation of key punching from source records. The most flexible machine which performs this as well as other functions is the Accumulating Reproducer. Reproducing in new cards need not be done in positions corresponding to those of the old cards. The positions can be selected and controlled by a removable and flexible control panel quickly inserted into the machine in much the same manner as in the Accounting Machine. This machine also will select and reproduce only those cards desired, without disturbing the arrangement of a file.

The **Accumulating Reproducer** will reproduce punching from one card into all cards following it until a new "master" card passes through the machine, instructing it to reproduce information appearing in that card. This is known as "gang" punching. These two forms of reproducing can be done simultaneously; that is, while information is being reproduced from one to another group of cards, additional data may be punched by means of one or more master cards. Both of these operations, working independently or simultaneously, take place at a speed of 6,000 cards per hour.

Other functions of the Reproducer include: printing in large size type, from data punched into the card, as many as eight figures on the edge of the card; comparing the reproduced and reproducing groups of cards to verify the accuracy of the reproduction; punching cards which have been



Fig. 3-38. IBM calculating punch, Type 602-A.

marked-sensed by a graphite pencil; and performing summary punching, as described previously. Summary punching and reproducing can be done simultaneously.

Calculating Punches. (Figure 3-38). To perform the routine as well as the complex calculations encountered in commercial and scientific work, IBM has five punched-card calculators, all varying in capacities and speeds.

One of these machines is the Calculating Punch. As cards pass through, this unit reads the factors, adds, subtracts, multiplies, divides, and punches the results. A multiplicand punched in the card can be multiplied by a factor in the same card or by a group multiplier punched in a single master card. There are several methods of controlling the machine to check the final results.

Several independent problems can be performed on this machine and several results punched in the card. The multiplier may contain as many as eight digits, the product as many as 30. Multiplying speed varies with the size of the factors and may be as high as 3,000 extensions an hour. The machine has a capacity for an 8 position divisor, a 15 position dividend, and an 8 position quotient. Dividing speed, as in multiplying, varies with the size of the factors to as high as 1,600 divisions an hour.

Various series of basic operations can be performed in any sequence as

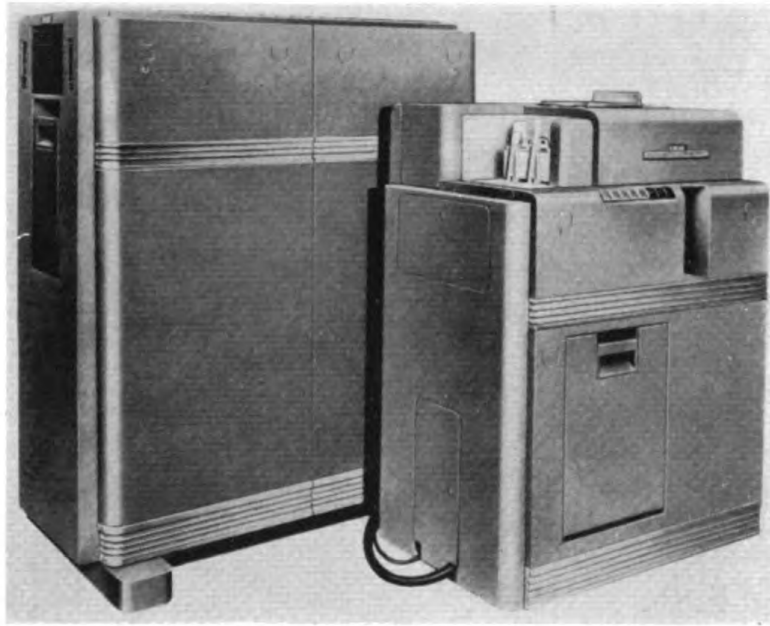


Figure 3-39. IBM electronic calculating punch, Type 604.

a card passes through the machine. Different factors may be added and the sum used as a multiplier, multiplicand, dividend, or divisor. A product or quotient can be further multiplied or divided by additional factors, and amounts may be added to or subtracted from the calculated results, the result being punched in the card in the same operation. Speed depends upon the complexity of the problem and the size of the factors.

The Electronic Calculator, Type 604 (Figure 3-39) is used when greater speed is essential. Calculations performed by this machine are made at a rate of 6000 cards per hour regardless of the operations involved. This machine functions in much the same way as the calculating punch described above. A point of difference is the fact that the calculating and punching operations are performed by different units. When even greater capacity and speed are important, the Type 607 Electronic Calculator may be used.

The Card-programmed Electronic Calculator. For problems requiring long series of arithmetical operations to obtain a single solution, the Electronic Card-programmed Calculator is widely used, since it performs these operations automatically. This ability, in addition to its storage capacity, high-speed computing, and high-speed printing, makes this calculator especially useful for the more complex problems encountered in engineering, statistical, and scientific work as well as in commercial applications.

A later development, the Magnetic Drum Data Processing Machine,

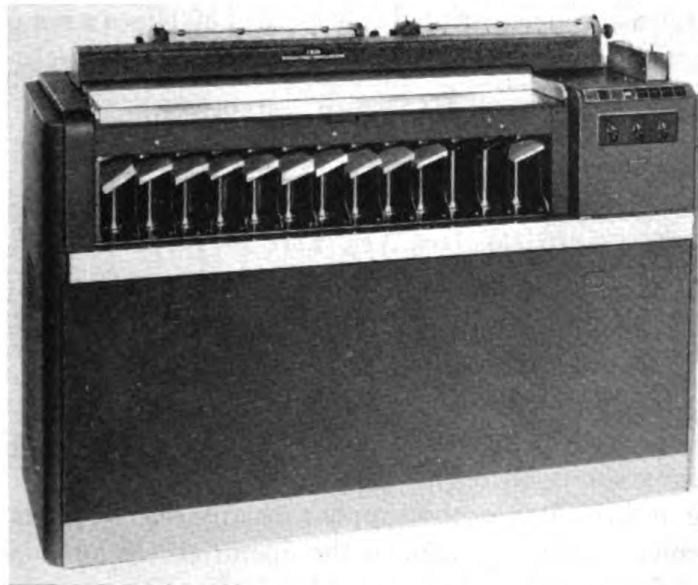


Figure 3-40. IBM electronic statistical machine, Type 101.

employs more advanced storage techniques that make it possible to process problems requiring internal storage of as many as 20,000 decimal digits.

The Electronic Statistical Machine (Figure 3-40) combines in a single unit the functions of sorting, counting, accumulating, balancing, editing, and printing of information. Unit counts may be distributed into as many as 60 different classifications while the basic data are sorted at the rate of 450 cards per minute in any desired sorting pattern to provide for further cross classifications. During the same run, information in the cards may be automatically checked on the basis of pre-established criteria for consistency. Files of cards may be searched automatically for specific numbers or ranges of numbers. By eliminating intermediate card handling and processing on other machines, the statistical machine provides a means for obtaining comprehensive statistical analyses in a relatively short time.

One of the most important developments in the field of high-speed data processing is IBM's Type 709 Data Processing Machine. Made up of varying numbers of units, depending on the work to be performed, the outstanding characteristics of the "709" are its very large storage capacities, and its very fast reading, writing, and computing speeds. The principal contributions to its speeds and to its flexibility in processing many types of problems are made by its three advanced types of storage—magnetic tapes, magnetic drums, and magnetic cores. As an example of the over-all speed of the "709", it is capable of performing up to 42,000 mathematical operations a second.

Input of data is by means of punched cards or by magnetic tapes. Out-

put is in the form of reports printed at a speed of 500 lines a minute, punched cards, or magnetic tapes.

The Remington Rand System

REMINGTON RAND, INC.

315 4th Ave., New York 10, N. Y.

The Remington Rand punched-card accounting method is based on punching holes in 90-column tabulating cards to code information pertinent to each business transaction or subject of interest. The holes in the card actuate machines that sort, add, subtract, multiply, punch, collate, interfile, and prepare in printed form records and reports in accordance with a company's over-all record keeping requirements.

In making an installation, the supplier usually selects persons employed by the customer and trains them in the operation of the equipment. The training time for each machine varies from a few hours to several days. The machines which have a keyboard like a typewriter demand the development of some skill, and require from two to three weeks for the operator to acquire a moderate amount of speed. However, several hours of instruction usually enable one to understand the operation of most machines.

Trained mechanics are located in major cities to service the equipment and to keep it in good operating condition. The machines are either sold outright or leased on an annual basis. When these are leased the rental includes service, but for machines sold outright there is available a mechanical service agreement under which, for an annual fee, Remington Rand agrees to keep them in operating condition. The purchaser may assign one of his own representatives to be trained in the maintenance of the machines at the supplier's mechanical training school. Parts are supplied at a nominal cost.

Tabulating cards (Figure 3-41) measure $3\frac{1}{4} \times 7\frac{3}{8}$ inches with a thickness no more than 0.007 inch or less than 0.00625 inch. These cards are manufactured to exacting specifications from a high quality paper stock and are printed to meet the needs of the specific application.

On the card illustrated there are 540 punching positions divided evenly among 90 vertical card columns, 45 on the upper half and the same number on the lower half. The punching code used for the six punching positions provides for recording 37 characters, 0 through 9, and A through Z, and one special character. Only one of the 37 characters may be coded per column, except under special conditions where each of the six positions in a column may be wired to a specific character.

Automatic punches (Figure 3-42) operate by simultaneous perforation on the punch die principle. This feature, exclusive with Remington Rand,

provides for setting of the dies in the punch by depressing the keys and for perforating the entire card with a single depression of a "trip" key. This enables the operator to correct an error before the card is punched. Also information common to each card may be recorded into a series of cards without resetting the keyboard for each card. Two basic key punches are available to cover the general range of manual punching requirements. One of these is for numerical punching only and the other for both numerical and alphabetical punching. Devices and attachments are available to modify these basic machines for particular applications. In addition, both machines may be used for "verify" punching as well as for the original punching entry. The production capacity of these machines varies with the design of the tabulating card and the type of information to be punched in each card. It is claimed that the average operator can punch 1500 90-column cards in an 8-hour day, but this may vary from a few hundred to as high as 5400 an hour depending upon the amount of variable information punched into each card.

Automatic Verifying Machine. To verify the correctness of punching in a pack of cards, a second operator repeats the original punching operation with a control on the automatic punch set for verify. This operation elongates all correctly punched holes and leaves round holes where there are errors. The cards are then taken to the Automatic Verifying Machine which mechanically senses the verify-punched cards. This machine places a card of contrasting color having uncut corners on top of each card containing a round hole. It also punches a small hole in the right-hand margin of each card passing through the machine to show that it has been verified. This machine has a constant production speed of 200 verify-punched tabulating cards per minute. For standard card-per-card interleaving the speed is 400 cards per minute.

The **Sorting Machine** embodying mechanical principles of reading the holes punched in the cards is known as the Standard Sorter. It operates automatically at a speed of 420 cards per minute. It requires only the insertion of the cards and their removal in sequence. The machine stops automatically when the last card is fed out of the feeding magazine, when a card becomes wrinkled or damaged, and when the receiving magazine or magazines are filled to capacity. One operator can handle up to four sorting machines depending upon the complexity of the operation. Additional devices are available for the Standard Sorter which permit group sorting (searching), and an operation which is known as "pairing," and counting.

The **Electronic Sorter**, (Figure 3-43) a new addition to the Remington Rand punched-card line employs the principle of "black light" reading of the card; it operates at the high speed of 800 cards per minute. Different from the standard sorter, this machine is equipped with 13 receiving maga-

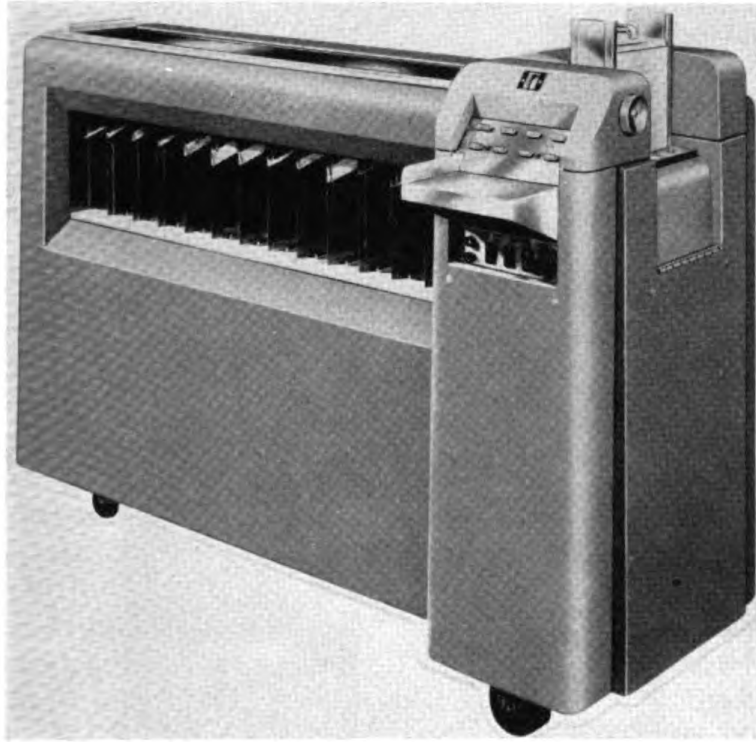


Figure 3-43. Remington Rand electronic sorter.

zines and one reject pocket which permits alphabetical sorting with one and one-half passes of the cards through the machine. In other words, on the first pass through the machine, the A-M letters are sorted in exactly that sequence and after clearing the receiving pocket, the remainder of the deck containing the N-Z letters are put through the machine and these letters are sorted in exactly that sequence. Any numerical punching which may occur during either of these passes will automatically fall in the reject pocket. The Electronic Sorter embodies all of the safety devices in the form of automatic stoppings that are included on the Standard Sorter described above.

Printing Tabulators (Figure 3-44) transpose the punched information from the card into a printed report. They are of two types—numerical and a combination numerical and alphabetical. All alphabetical tabulators have type bars which print any numerical figure from 0 through 9 and letters from A through Z and one special character. The large capacity alphabetical tabulators have up to 100 sectors and will print 100 characters simultaneously. Counters for the addition or subtraction of numerical information can be installed on all alphabetical tabulators. Two counters, one to give totals and the other grand totals, can be attached to each type bar. Each of 80 type bars can be equipped with two of the direct subtrac-



Figure 3-44. Remington Rand alphabetical tabulator.

tion type counters, thus providing 160 counters for adding and subtracting. Output is at the rate of 6,000 machine cycles per hour. A machine cycle consists of printing and accumulating one tabulating card or printing one total from a group of tabulating cards.

The **Punched-card Interpreter** prints across the face of a tabulating card whatever alphabetical and numerical information is punched into the card. The location of each printed character is normally directly above the zero position in the column in which it is punched, but, if desired, it may be placed on any one of thirteen printing positions on the upper half of the card, at a speed of 90 cards per minute.

The **Posting Interpreter** makes possible taking information which is punched in one card and printing it on a following card. Two important applications are the preparation of tabulating card checks and the posting of employees' earning records on tabulating card ledgers from information punched in weekly net earnings summary cards.

The **Posting Machine** is the automatic line finding posting machine which permits the use of tabulating cards as ledgers for historical, chronological, or continuous listings of transactions that occur within an account. Prior to the posting operation, the punched cards created either by key punching or summary punching are collated with the ledger cards and the posting machine reads the accounting information punched in the leading card, automatically selects the next open line for posting on the ledger card and prints the information on the ledger. This posting machine operates at a speed of 90 cycles per minute which gives a net productive posting speed of 45 postings per minute. It is equipped with a dual-card receiving magazine so that the punched detail cards are automatically segregated from the printed ledger cards without going through a subsequent machine operation to separate the two sets of cards.

The **Multi-control Reproducing Punch** reproduces information from one tabulating card into an unpunched card. Its function is to compare two cards with respect to their coding and then to reproduce the variable data from the primary (master) card into one or more secondary (detail) cards, depending upon the existence of a matching or non-matching condition. At the same time cards that match can be segregated from cards that do not match. The output of this machine is from 100 to 200 cards per minute depending upon the operation being performed.

The **Interfiling Reproducing Punch**, an improvement over the Multi-control Reproducing Punch, is used for comparing two decks of cards and for punching, segregating, or interfiling, depending upon the results of the comparison. Its speed is identical to that of the Multi-control Reproducing Punch.

The **Collating Reproducer**, (Figure 3-45) a modification of the Interfiling Reproducing Punch, automatically compares, punches, interfiles, segregates and sequences 90-column tabulating cards. Thus, it can be used to verify, group-extend, code, decode, in-file, and out-file. Any errors found by this machine when checking the sequence in a single card file are marked by signal cards and the erroneous cards can either remain in the file or be removed. Two separate card files arranged in numerical sequence can be interfiled. This machine also controls the feeding of cards from one or more feeding magazines in numerical sequence. Segregation and selective punching can be accomplished simultaneously. These operations are performed at rates of 6,000 to 12,000 cards per hour.

The **High-Speed Electric Collator** (Figure 3-46), operating at a speed of 250 machine-cycles per minute, embodies the electric principle of sensing cards. The collation is controlled by a flexible wiring panel which can be readily changed by the operator; or extra pre-wired panels are available which can be readily installed to change the setup. A feature of the Elec-



Figure 3-45. Remington Rand collating reproducer.

tric Collator is its ability to check the sequence of both sets of cards being fed from the two feeding magazines without loss of its ability to merge or segregate in the four receiving pockets. Although the machine is basically a numerical unit, it nevertheless will compare alphabetical information punched in the two files and will segregate cards which do not match.

The **Calculating Punch** performs the arithmetical operation of addition, subtraction, multiplication and division from values sensed from punched cards. The results of these operations are then punched into the same card from which the values were obtained or into any card which follows.

In arriving at the final result or results for an application, the machine follows a pre-planned course of operation for the processing of one card through the machine. This is called a program. The individual elements of the program for any given application are program steps. Provision is made for 12 program steps. In any one application, part or all of the steps may be followed as required. It is also possible to expand the program by repeating steps. Within one program, two or more subroutines (sub-program) involving one or more steps may be followed. The Calculating Punch will also add or subtract two or more values sensed from the same card to arrive at one or more results. The result, or results, thus obtained

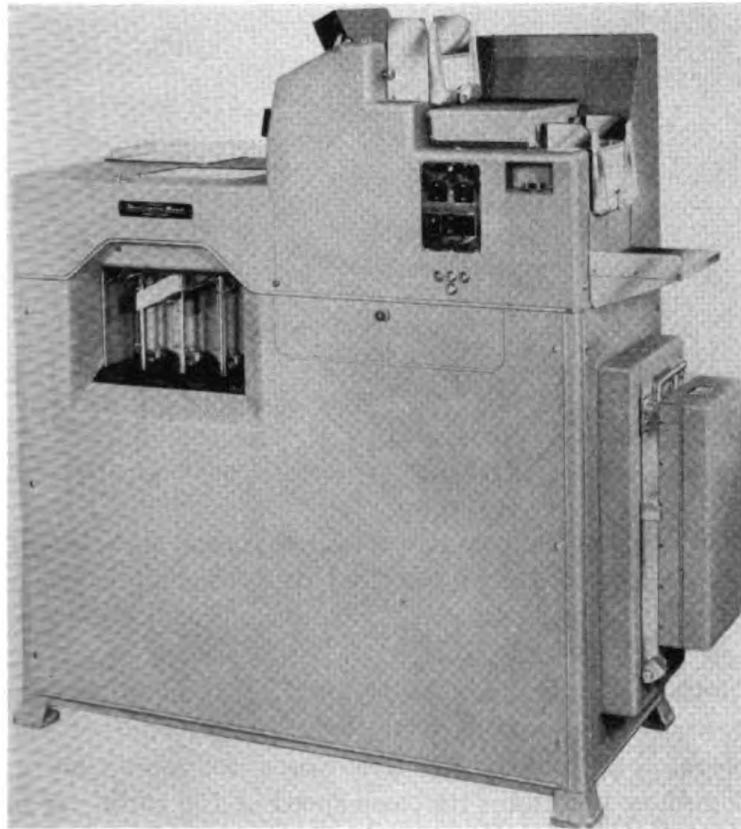


Figure 3-46. Remington Rand electric collator.

may be used for additional calculations for the same card or the following card. This is called "cross-footing."

A feature of the calculating punch is that values arrived at early in the program for one card may be punched into that same card while the machine proceeds with further calculations for the same card. On the other hand, the machine will also proceed with the calculations for one card while simultaneously punching the result into the card immediately preceding it.

The entire machine performance for individual applications is obtained through, and is controlled by, the panel. By means of this panel all phases of machine operation may be varied from application to application to meet the individual requirements. The operator may pre-wire this panel as required for individual application, or he may readily install a pre-wired panel for the new or subsequent application.

The **Punched Card Electric Computer** (Figure 3-47) performs the arithmetical functions of addition, subtraction, multiplication and division of values sensed from punched cards, values manually set into the machine, or values computed during a sequence of programming a problem. Results

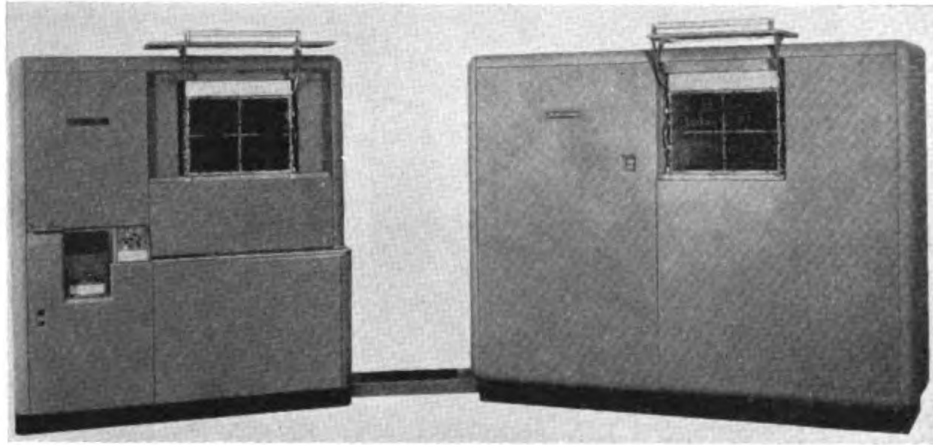


Figure 3-47. Remington Rand punched card electronic computer.

of these functions are then punched into the same card from which the values were sensed, or into any desired card which follows, or they are placed in storage for subsequent use in the computations or over-all accumulations.

The Computer operates at a basic speed of 150 cards per minute but will not proceed to the next card until the computation of the card then being processed is completed. In other words, for most commercial problems the Computer maintains its basic speed of 150 cards per minute, but on long iterative problems the output speed will be lowered in order that the Computer may complete the computation.

The Computer has provision for 40 program steps. The programming is accomplished by means of a flexible wiring panel.

Features of the Remington Rand Punched Card Computer are: (1) each program step is self-checking without requiring another program step; (2) the program steps may be used as often as required by means of "branching" or "selectors"; (3) alphabetical data may be transferred from "master" cards to following "detail" cards; (4) the dual card receiving magazine permits the segregation of "master" cards from "detail" cards or negative from positive results, etc.

The Computer is also capable of solving problems in higher mathematics thereby making it suitable for engineering, scientific and research programs.

"Synchro-matic." This electrical synchronization of a tabulating card punch and a Remington bookkeeping machine automatically punches information into tabulating cards simultaneously with the recording of that information on an original record by means of the accounting machine. The latter controls the rate of the operation.

The **Summary Card Punch** (Figure 3-48) punches alphabetical characters and numerical information, both designating and adding, into a

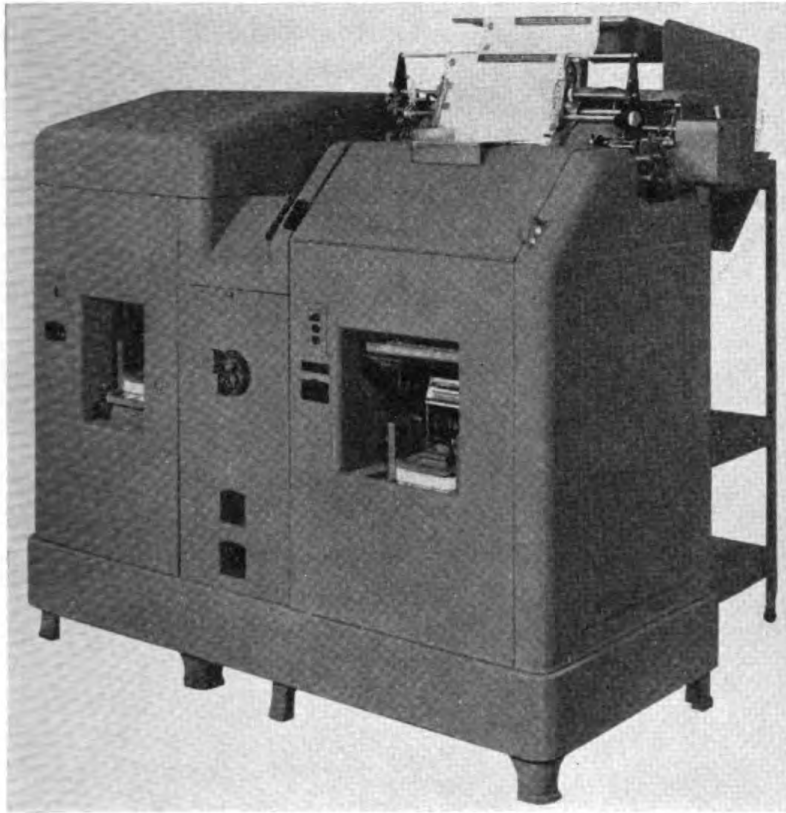


Fig. 3-48. Remington Rand summary card punch, combined with an alphabetic tabulator.

tabulating card automatically and simultaneously with the printing of the information as a group total by the tabulator. The capacity of this machine is determined by the number of group totals produced by the tabulator with which it is connected.

Tag Control Reproducer is a tabulating card punch which reads "pin holes" punched in garment price tags and in turn, at a speed of 100 tags per minute, punches the recorded data such as vendor, style, size, color, selling period (season), store number or department, price and cost in standard tabulating cards.

These cards are then used to prepare reports of sales by the various codes mentioned above, as well as inventory balance reports, thereby providing management with up-to-the-minute reports for effective merchandising.

Tape-to-Card (Figure 3-49) and **Card-to-Tape Converters** are also being used in industry and transportation. Data typed on communication equipment such as "teletype" is also recorded by means of holes punched in paper tape. This tape in turn activates the tape-to-card unit which punches the data in tabulating cards.

Similarly, data recorded in punched cards, which are then put through



Figure 3-49. Remington Rand tape-to-card converter.

the Card-to-Tape Converter, results in a punched paper tape which may then be transmitted automatically over wire communication systems to remote points. This equipment was originally developed for the transportation industry, but industry in general has found many uses for the same machines and is presently using it to tie in remote warehouses with central accounting and billing points.

SUPPLEMENTS TO PUNCHED CARDS

Filmsort

FILMSORT DIVISION, DEXTER FOLDER COMPANY

Pearl River, N. Y.

The Filmsort system converts microfilm into a punched-card tool. Reference and research materials of varying length and size are copied on 16- or 35-mm microfilm with standard cameras. The individual strips or frames are fitted into standard, uniform Filmsort cards, which may be plain index cards, edge-notched punch cards, or standard tabulating cards.

After microfilm insertion, the punched cards may be processed through

any of the standard punched-card units. The use of this system results in space-saving, security and faster reference.

Furthermore, Filmsort makes good use of microfilm as an inexpensive copying medium. Reference material may be copied and the film filed in card form for bibliographies.

Filmsort Aperture Card (Figure 3-50). Individual frames of 16-35-mm microfilm are housed in the Filmsort aperture card. An aperture card is simply a standard punched card or index card converted for microfilm

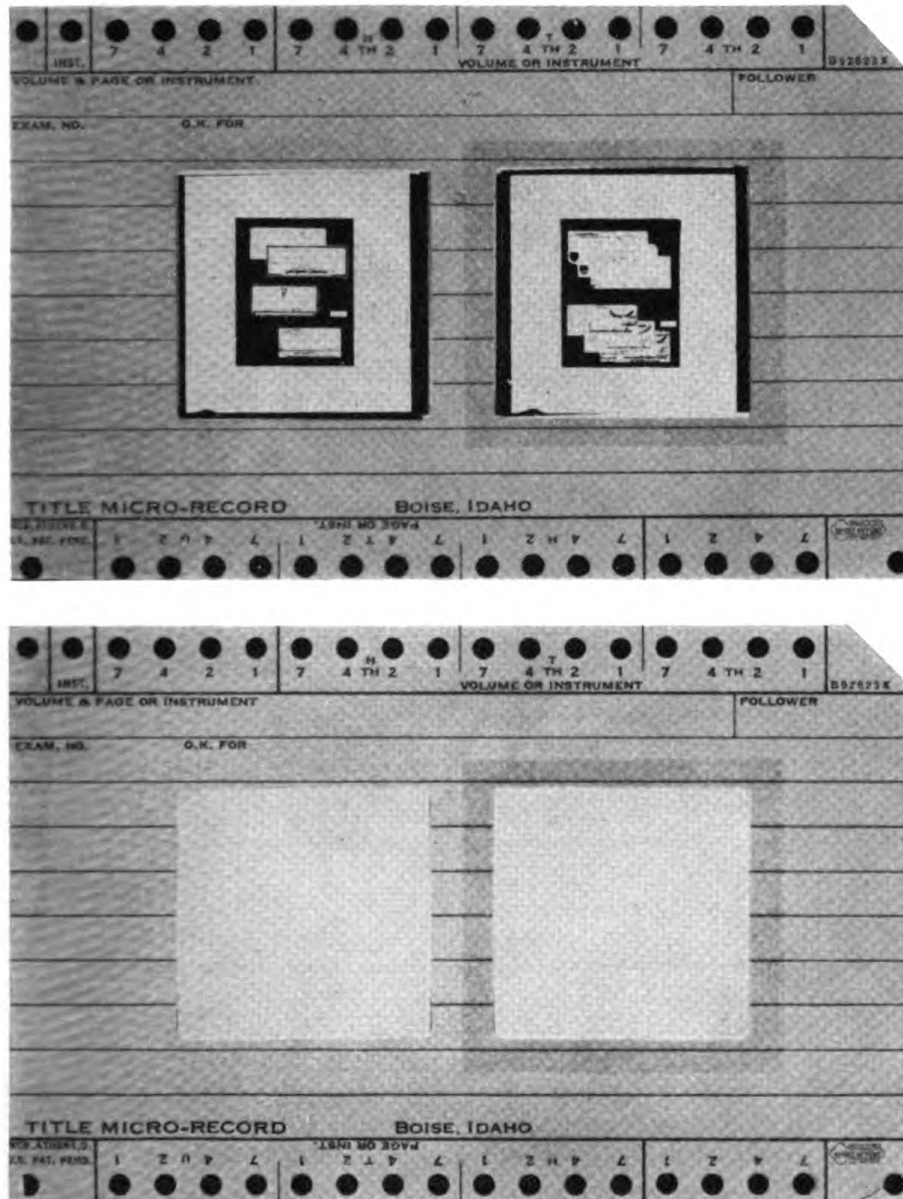


Figure 3-50. Application of a Filmsort insert to a "Keysort" card.

insertion by Filmsort. Filmsort takes the cards which the customer has selected, die cuts a window and applies a pressure-sensitive adhesive that holds the microfilm. A glassine sheet is put into the window to protect the adhesive until the microfilm is inserted into place.

After the film is inserted into the card, the punched card with microfilm may be sorted, selected, collated and handled in any of the standard punched-card procedures.

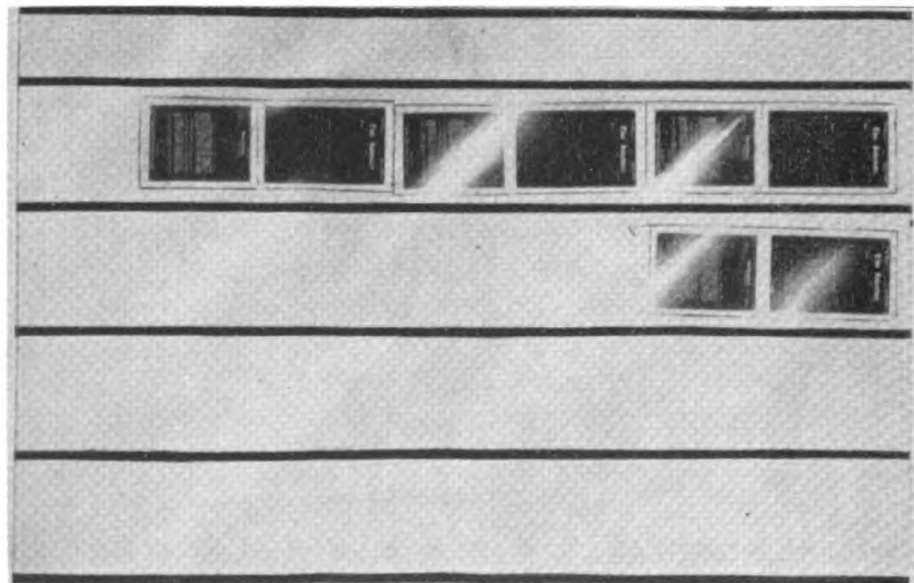
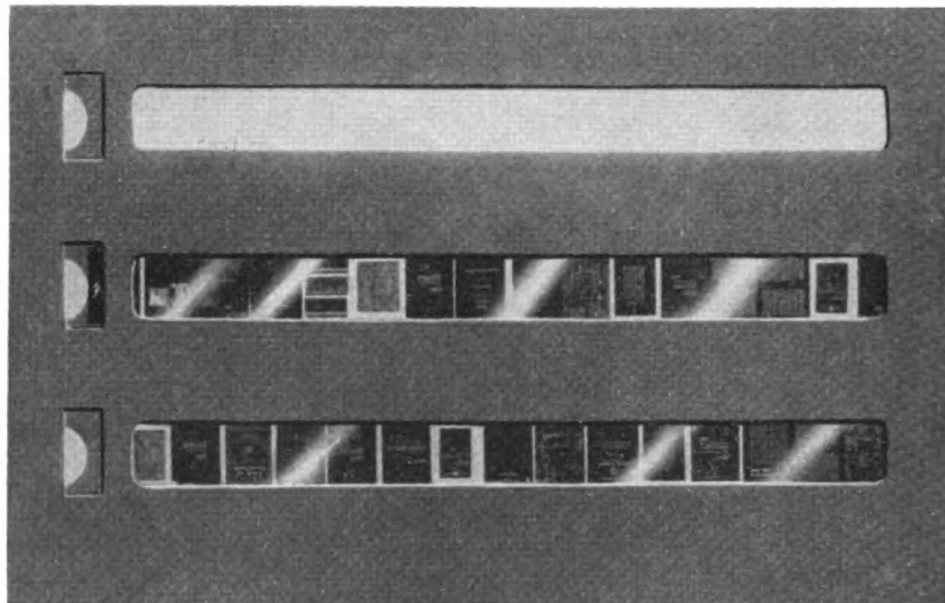


Figure 3-51. Filmsort jacket cards, paper and acetate.

Filmsort Jacket Card (Figure 3-51). Developed for the multi-paged record of varying length, the Filmsort jacket card condenses a file folder of data to standard card size. Strips of microfilm are inserted into the slot between the two layers of the jacket and the microfilm can be read without removing it from its jacket.

Jackets are manufactured in acetate or paper models, in standard 3 x 5, 4 x 6, and 5 x 8 inch sizes. In cooperation with Royal McBee, any Filmsort paper jacket can be processed for McBee form printing and Keysort marginal punching techniques.

Jacket cards condense the contents of 12 to 14 file cabinets into one cabinet of Filmsort cards. The acetate jacket stores the most film in the least space. A report of 120 pages, each 8½ x 11 inches, becomes as small as one Filmsort 5 x 8 inch jacket.

Filmsort Mounter (Figure 3-52). More than 350 individual microfilm frames may be mounted hourly into Filmsort aperture cards. In one stroke, the mounter cuts the frame of the film from the microfilm reel and simultaneously attaches it to the Filmsort card. The rate of production for the mounter compares favorably with that of standard planetary cameras. A new automatic machine (Figure 3-53) for mounting film into cards at the rate of 2,000 cards per hour has recently been developed.

Filmsnips (Figure 3-54). This tool is a hand-operated scissors-like die used to cut out microfilm for manual insertion into Filmsort aperture cards. It is recommended for those who mount 50 frames of film or less per day.

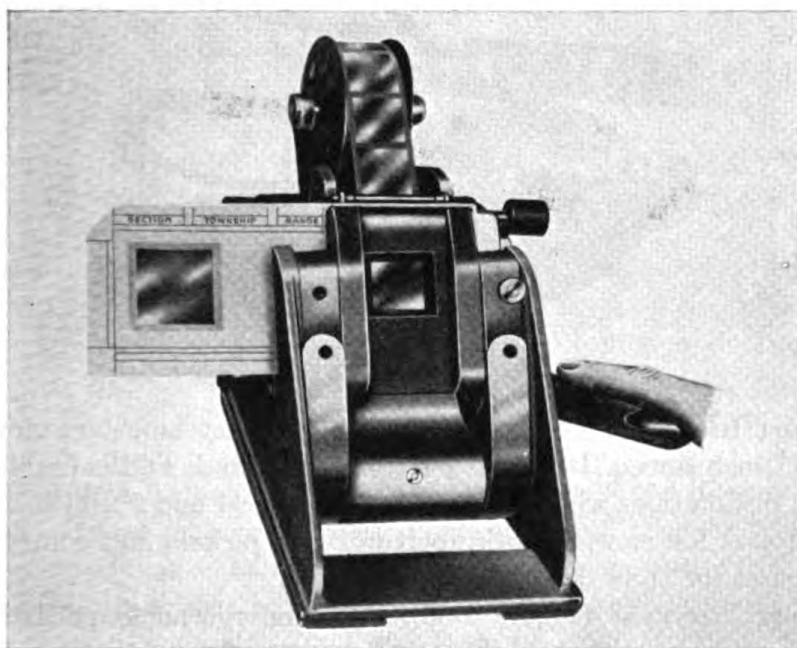


Figure 3-52. Filmsort mounter.

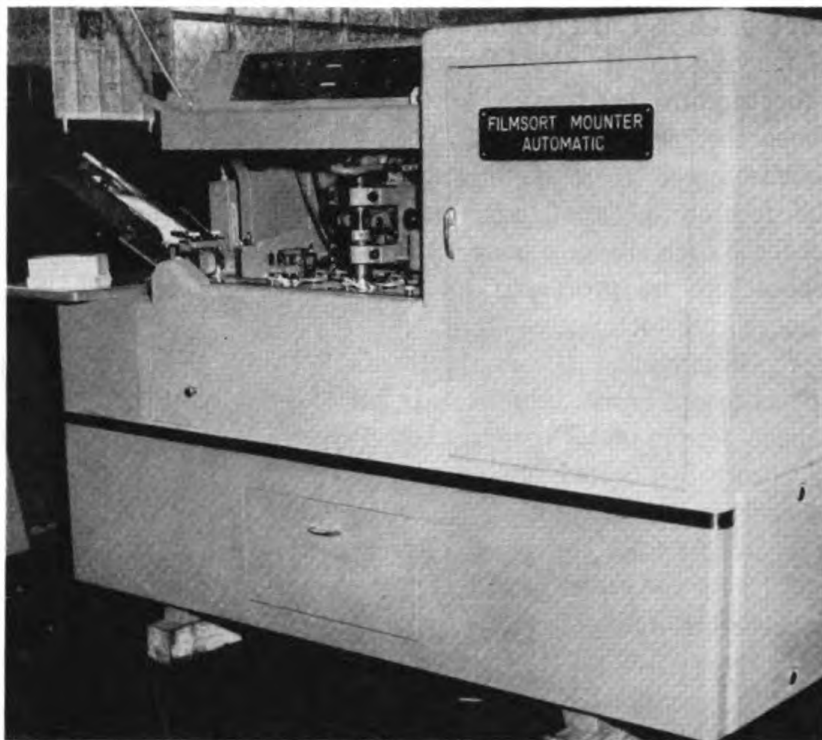


Figure 3-53. Filmsort automatic mounter.

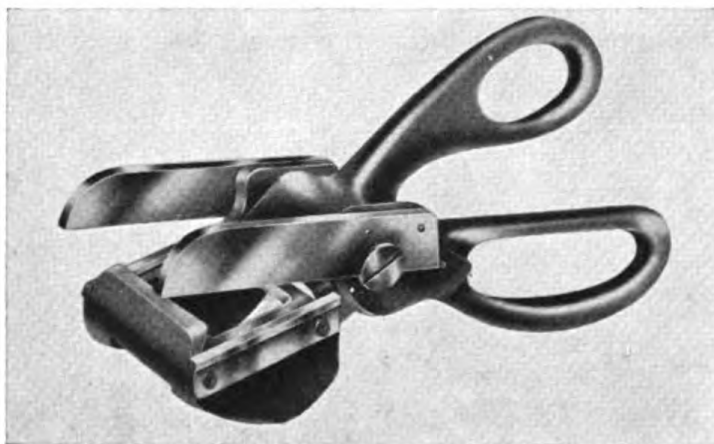


Figure 3-54. Filmsnips.

Filmsort Inspector (Figure 3-55). This compact table top viewer has an 11 x 11 inch screen. It weighs less than 20 pounds and is usable in all microfilm installations where copy is in the 8½ x 11 and 9 x 14 inch range. It is equipped for viewing both apertures and jackets and comes in two different sizes for 16 or 22 times magnification.

Filmsort Surveyor (Figure 3-56). Built for viewing large-sized copy, the Filmsort Surveyor comes in two models—18 x 24 or 24 x 36 inch screen



Figure 3-55. Filmsort Inspector.

size. It is equipped with variable magnification and has an enlargement ratio that can be adjusted from 10.5 to 22 times.

Magnification is automatic by throwing a switch which starts a chain drive that increases or decreases the image. The Surveyor covers a full $1\frac{1}{4} \times 15\frac{1}{8}$ inch microfilm frame.

Filmsort Reviewer (Figure 3-57). Where a high intensity light source is needed for viewing microfilm, the Filmsort Reviewer is applicable. Its optical system and reflectionized screen give faithful reproduction of such complicated microfilmed materials as radiographs, charts, graphs, etc. Because of its high light source, the Reviewer can be employed as a microfilm enlarger, using silver chloride papers.

Reading and Enlarging. All film is read or enlarged directly from the Filmsort aperture or jacket. Most standard microfilm enlargers will take any Filmsort jacket and the majority of Filmsort aperture cards.

Reproducing Aperture Cards. Two methods have been developed for making positive microfilm duplicates from an original Filmsort aperture



Figure 3-56. Filmsort Surveyor.



Figure 3-57. Filmsort Reviewer.

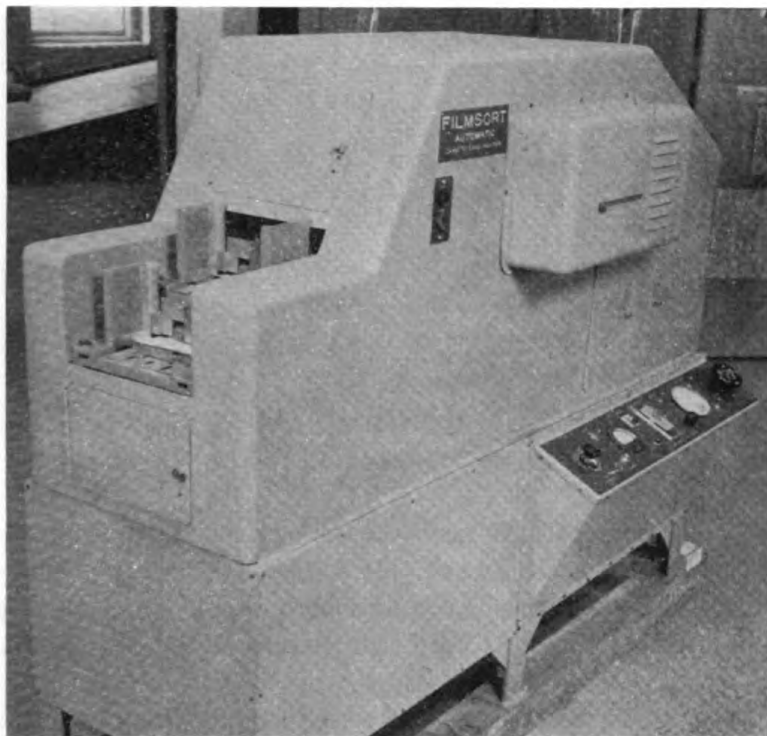


Figure 3-58. Filmsort card-to-card duplicator.

card. The Kalfax process uses an ultraviolet-sensitive film which is developed and fixed by heat. The Ozalid process uses a card with an unexposed ozalid duplicating microfilm in the aperture. It is processed by the standard diazo process.

A recent development is the **Card-to-Card Duplicator** (Figure 3-58) which is capable of reproducing Filmsort cards complete with mounted image at a rate of 2,000 an hour.

Distribution. Remington Rand, the Recordak Corporation, the Ozalid Division of General Analine and Film Corp., and Microdealers, Inc. are the national distributors for the various Filmsort products.

Microtape

THE MICROCARD CORPORATION

West Salem, Wis.

Microtape (Figure 3-59) was developed by The Microcard Corporation as an economically feasible substitute for Microcards, where less than five copies of a document were needed. Microtape consists of 100-foot rolls of 16-mm or 35-mm positive microtext, printed from standard negative microfilm rolls, and having a pressure-sensitive adhesive laminated to the back.

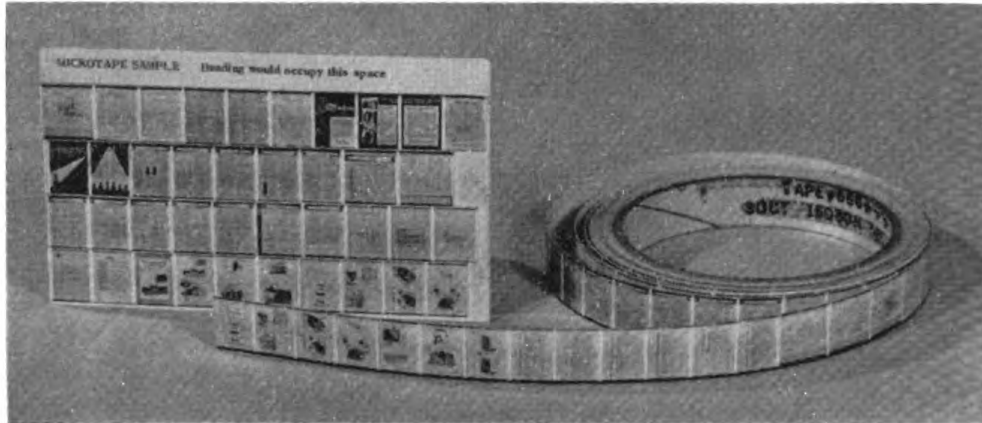


Figure 3-59. Microtape, in reel form and applied to card.

The user cuts apart the strips, peels off the protective backing, and presses the Microtape onto the filing card. Microtape can be applied to standard index cards, visible-index cards, and edge-notched cards. It cannot be used with tabulating punched cards at the moment because of its added thickness.

Three readers are available, two of which are desk models. One of these has a card-moving mechanism for ease in locating pages. The third is a portable pocket model with a battery and 110 volt sources of illumination.

The American Microfilming Service Company, Microtape Systems, of New Haven, Connecticut, has set up, under license from The Microcard Corporation, a chain of processors who will either arrange for the production of Microtapes from existing microfilms or take the microfilms and have the Microtapes manufactured.

OTHER TYPES OF INFORMATION SEARCHING DEVICES

Uniterm, Matrex and Radex System

DOCUMENTATION, INC.

2521 Connecticut Ave. NW., Washington 8, D. C.

The Uniterm System of coordinate indexing is a manual information retrieval system. It provides a card for each indexing term used to characterize the documents in a collection.

Each of these **Uniterm cards** is divided into columns headed by the digits "0" through "9". The number of a report to be indexed by a given term is entered on the term card in one of ten columns. The digit at the head of the column corresponds to the last digit of the report number. The term cards are filed alphabetically.

The collection is searched by comparing the term cards corresponding to

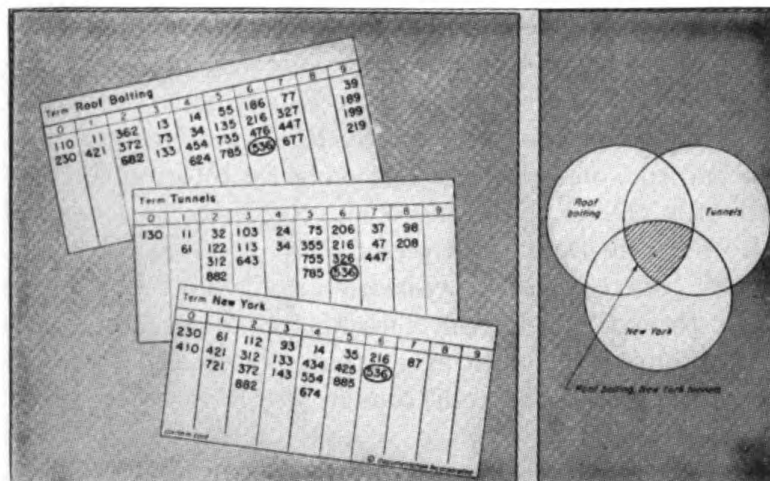


Figure 3-60. Uniterm card system.

the search question, for common numbers. These constitute the search results. An application of the Uniterm System is described in Chapter 7.

The Uniterm System comes in two forms. The Uniterm-Card system, illustrated in Figure 3-60, is used where only a single index is required. Uniterm cards and instruction books are available from Documentation, Inc.

Where the need exists for a low-cost large scale dissemination of an index, the Uniterm-Book system illustrated in Figure 3-61 can be used. Each

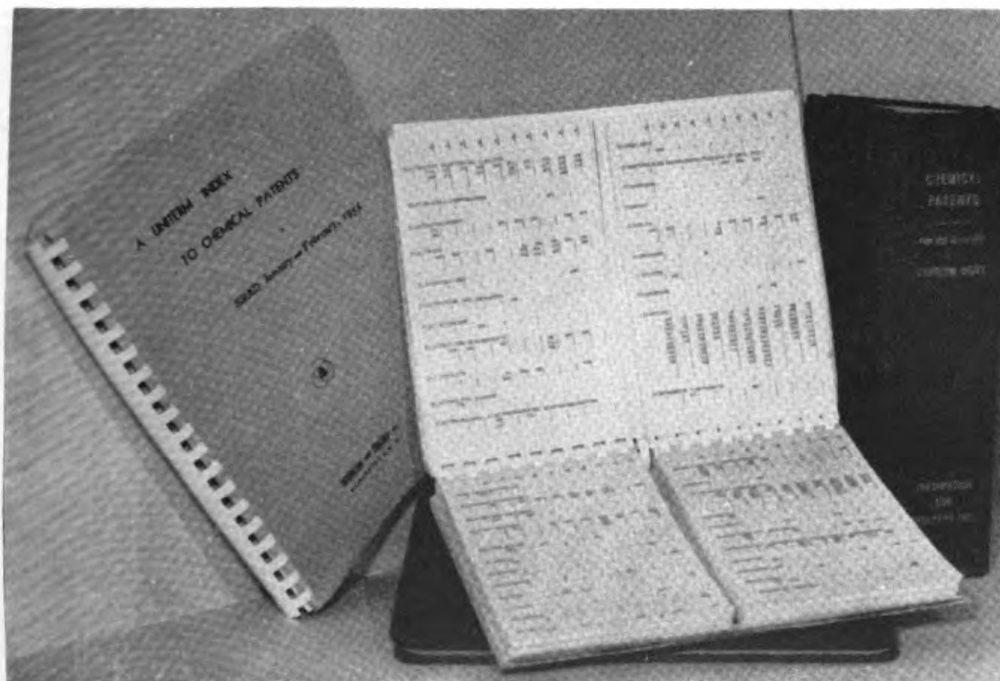


Figure 3-61. Uniterm book system.

page carries the reproduction of a number of Uniterm cards. In order to facilitate the comparison of terms for common numbers, two duplicate sets are provided within a single binding. This eliminates the need to turn pages in making comparisons. Uniterm-Book systems are prepared by Documentation, Inc. on a contract service basis. Information for Industry, Inc., Washington, D. C., publishes a current index to Chemical Patents based on the Uniterm-Book principles.

The **Matrex (matrix-index) systems** are machine systems specifically designed for information retrieval. Two Matrex systems now available are the Termatrex and Alpha-Matrex systems. They are based upon cards each representing the entire collection in the form of a matrix, as first proposed by Batten.

Termatrex systems feature a card for each indexing term used to characterize the item of information in a collection. Each term card has a small area dedicated to each document in the collection.

An item of information is entered into the Termatrex system by placing all term cards corresponding to the indexing terms by which that document has been indexed, in superimposition in the Termatrex device. A hole is then "punched" with a high-speed drill in all of these cards simultaneously.

The collection is searched by placing the term cards corresponding to the search question in superimposition in the same Termatrex device. The search results, in the form of coinciding holes in the question cards, are visible as light dots on a screen. The serial numbers of these items of information are read off directly.

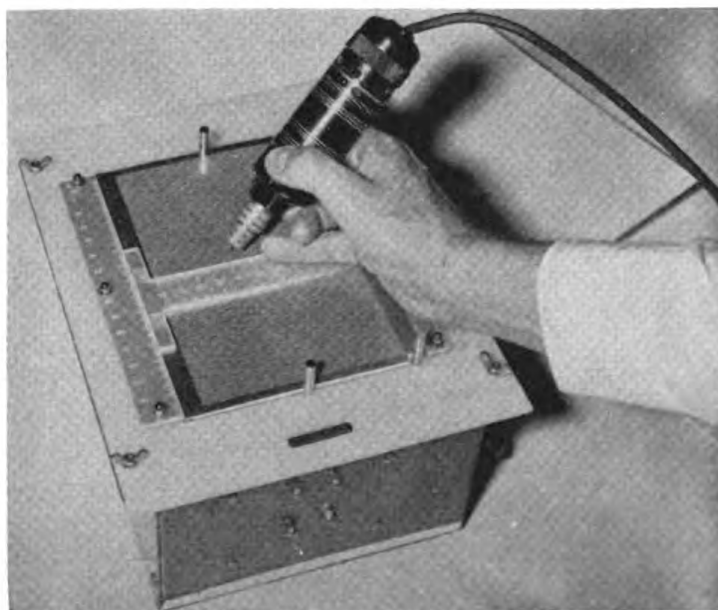


Figure 3-62. Termatrex-15 device.

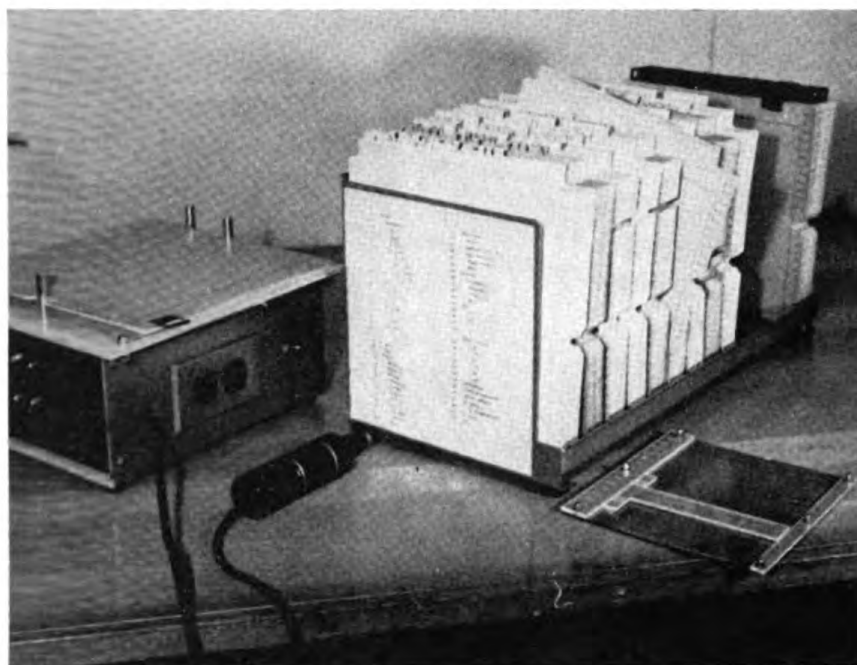


Figure 3-63. Termatrix-10 device.

Termatrix devices come in three models:

Termatrix-15, shown in Figure 3-62, has a capacity of 15,000 items of information per set of 5 x 8 inch cards. It is intended for private collections where low cost is a first requirement.

Termatrix-10, shown in Figure 3-63, has a capacity of 10,000 items of information per set of 10 x 10 inch cards. It features wider spacing and larger holes, and is used for smaller library and industrial applications where ease of operation is a first requirement.

Termatrix-40, shown in Figure 3-64, is the all-purpose standard Matrex machine. It has a capacity of 40,000 items of information per set of $17\frac{1}{2}$ x $17\frac{1}{2}$ inch cards.

For applications where the data input load or the search load is extremely high, **Alpha-Matrex systems** featuring a higher degree of mechanization, are available.

In the Alpha-Matrex systems, cards are provided for each letter of the alphabet rather than for each indexing term. This results in a relatively small and fixed number of cards, and in a high storage efficiency. From seven to ten alphabets are used to avoid spelling ambiguity, (first letter alphabet, second letter alphabet, etc.).

In entering a new item of information in the system, the indexing terms used to index the item are typed out on a keyboard which actuates the Alpha-Matrex selector to deliver the corresponding cards. The selected cards are placed in the Termatrix-40 and the item is entered by "punch-

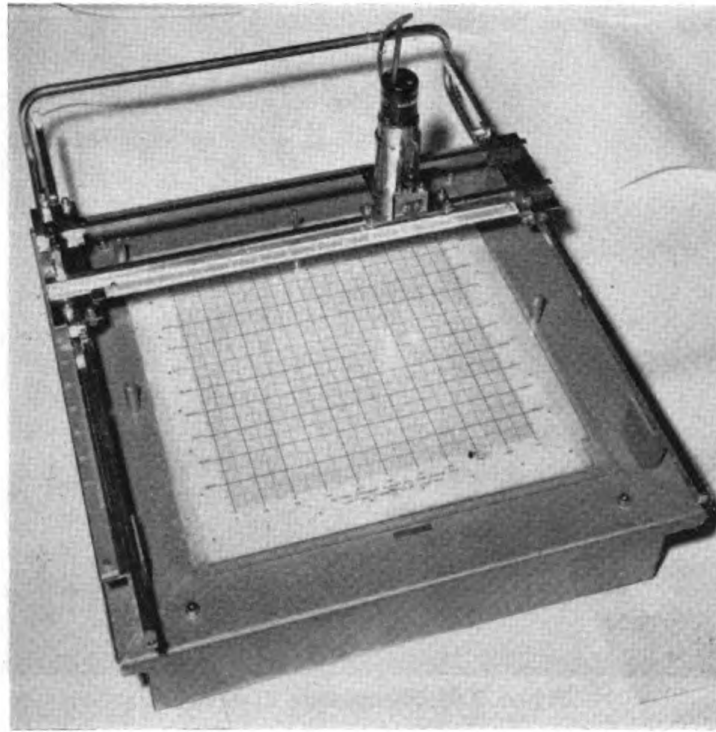


Figure 3-64. Termatex-40 device.

ing" a hole in all cards simultaneously at the position corresponding to the serial number of the item.

For searching, a similar selection procedure is followed. The selected cards are placed in the Termatex-40, and the serial numbers of light dots visible on the display screen are read out.

The Alpha-Matrex comes in two models. The mechanical Alpha-Matrex, shown in Figure 3-65, features an adding-machine type of keyboard.

The electrical Alpha-Matrex system, shown in Figure 3-66, features an electric typewriter which also provides a record of the typed terms for verification purposes. Errors can be corrected by means of correction buttons on the selector. After all terms have been typed out, the corresponding cards are instantaneously selected. These are then placed in the Termatex-40 machine. The capacity of the Alpha-Matrex systems is 40,000 items per set of cards.

Matrex systems can be used with any type of indexing system, whether it be based on Uniterms, Descriptors, a Subject Heading system or a classification schedule. Existing collections based on a Subject Heading or classification schedule can be entered into a Matrex System without re-indexing. This form of mechanization combines the retrieval possibilities of the Subject Heading or classification index with the retrieval possibilities provided by Uniterm or Descriptor indexing.



Figure 3-65. Mechanical Alpha-Matrix device.

Matrex systems can be provided with simple photographic equipment for photo-logical operations. This equipment allows a complete program comprising logical sums, products and complements, as well as generic searches, to be performed on the entire information collection in one sequence and in a matter of minutes.

The Mega-Matrix system, which is built to customer specifications, is capable of handling collections running into millions of items.

The **Radex filing system** is a novel numerical filing system, applicable to cards 5 x 8 inches and larger. The cards are grouped in decks of 100 and each card is provided with a small tab on which its terminal digit has been printed. There are a hundred different tab positions corresponding to each of the hundred different cards contained in the deck.

This system allows instantaneous identification of each card within a deck of 100. It likewise allows random refiling within each deck. Figure 3-63 shows Termatrix cards equipped with Radex tabs.

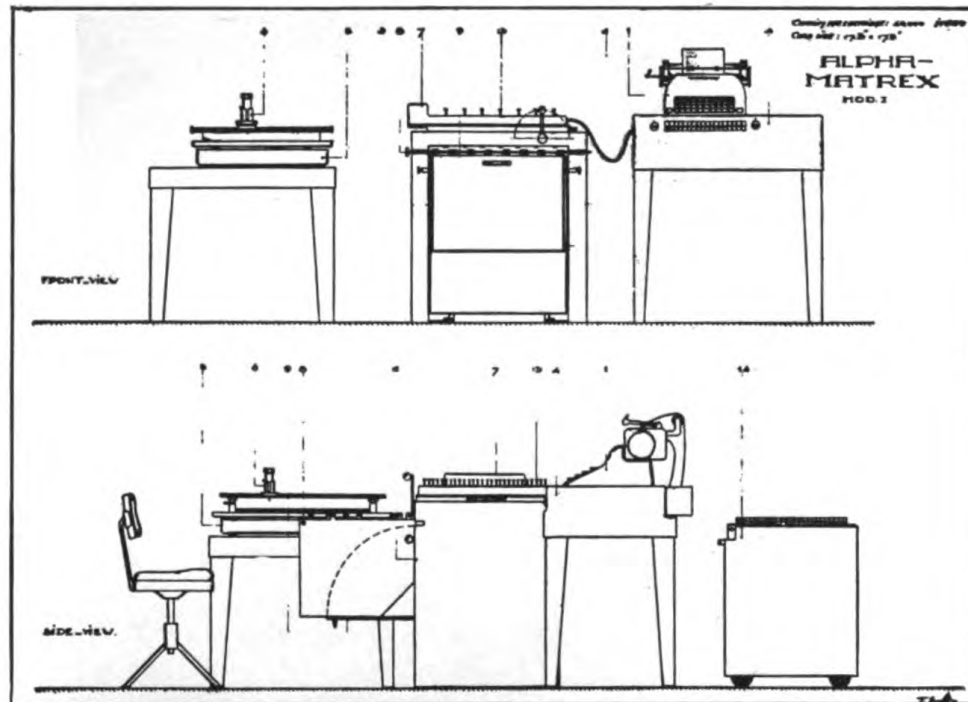


Figure 3-66. Electrical Alpha-Matrex system.

The Filmorex System

74, Rue des Saints-Peres

Paris 7^e, France

The Filmorex system operates with pieces of **film** (Figure 3-67) coded by means of light and dark spots, in much the same way that tabulating cards are internally punched. Each piece of film, 70 x 45 millimeters, is divided into two sections. One is a microphoto of the document and the other is the light and dark pattern of appropriate codes used for searching.

A microfilm camera is available with provision for producing the proper code patterns by means of a keyboard. The other main piece of equipment is the **selector** (Figure 3-68), which will read and select film at the rate of 36,000 pieces per hour.

Magnetic Inks, Punched Paper Tape, Magnetic Tape, Etc.

In the August, 1956, issue of *Banking* there appeared a report of a subcommittee of the American Bankers Association which investigated the mechanization of check handling by means of magnetic and luminous inks. About ten companies are listed there as participating in the research and development of these methods.

The June, 1957, issue of *Computers and Automation* contains an ex-



Figure 3-67. Filmorex film.



Figure 3-68. Filmorex selector.

haustive list of manufacturers of computers, components, magnetic and paper tape readers and writers, and other information processing equipment.

An article appeared in *Science* (October 26, 1956) by Karl Heumann on the use of computers and related devices in the field of information retrieval. The reader is referred to the above mentioned sources for further information on non-punched card equipment in use, or for possible use, in the storage and retrieval of information.

ANCILLARY EQUIPMENT

Many companies manufacture ancillary equipment such as card files, key punch desks, plugboard storage racks, index cards, and the like. Some of them are listed below. Again, this is not to be considered an exhaustive list.

Art Steel Company, Inc.
170 West 233rd St.
New York 63, N. Y.

Monarch Metal Products, Inc.
724 South Columbus Ave.
Mount Vernon, N. Y.

Berger Division
Republic Steel Corporation
E. 11th Street & Belden Ave.
Canton 5, Ohio

Record Files, Inc.
1490 Lincoln Highway West
Wooster, Ohio

Duro Consolidated, Inc.
P. O. Box 248-1
Redwood City, Calif.

Shaw-Walker
1950 Townsend St.
Muskegon 6, Mich.

Dresser Products, Inc.
152 Wheeler Ave.
Providence 5, R. I.

Tab Products Company
57 Post St.
San Francisco 4, Calif.

Globe-Wernicke Company
5029 Carthage Ave.
Cincinnati, Ohio

The Wright Line, Inc.
100 Exchange St.
Worcester 8, Mass.

Acknowledgment

Mr. Harry C. Zeisig, Jr., of the Spencer Chemical Company, originally assembled much of the information for this chapter. Due to the pressure of other business he was unable to complete the collection of material. The author wishes to acknowledge Mr. Zeisig's important contribution to this chapter.

Part II

PRACTICAL APPLICATIONS OF PUNCHED CARDS AND RELATED DEVICES

Chapter 4

THE USE OF PUNCHED-CARD TECHNIQUES IN PREPARING REPORTS, PAPERS, AND BOOKS*

CHARLES A. BURKHARD

Locomotive and Car Equipment Department, General Electric Co., Erie, Pennsylvania

Introduction

A research project is not finished until the results have been reduced to writing, in the form of a report, a paper, or even a book. Report writing and manuscript preparation have been widely discussed. This chapter will deal only with the mechanical aspects of the problem.

In general, an outline is prepared and pertinent data are then inserted in their logical places. This step involves selecting, sorting, arranging, and correlating large masses of data. Other chores connected with technical writing, such as the arrangement of the bibliography, also involve much tedious shuffling of papers or cards. If conventional methods are used, these tasks, the performing of purely mechanical operations, may take an undue proportion of the writer's time. This chapter will show how punched-card techniques can be used as an aid to technical writing by facilitating the time-consuming mechanical operations. To illustrate the effectiveness of such techniques an account is given of the use of punched cards in preparing a review paper on organosilicon chemistry.¹

Description of Punched-Card File†

Author, title, reference, and abstract, including physical and chemical data, as well as other pertinent bits of information, were written or typed on the cards. In general, all information of interest found in each paper,

* This chapter is based on a paper, entitled "The Use of Hand-Sorted Cards in Small Files", presented before the Division of Chemical Education of the American Chemical Society at the 113th national meeting in Chicago, Ill., April, 1948. Thanks are due the American Chemical Society for approving publication of the revised paper in this book. The author wishes to acknowledge the valuable help given by Dr. A. E. Newkirk, Research Laboratory, General Electric Co., Schenectady, N. Y., and by others.

¹ Burkhard, Rochow, Booth, and Hartt, *Chem. Rev.*, **41**, 97 (1947).

† Five by eight-inch McBee "Keysort" cards were used (Figure 4-1).

publication, or company report was entered on a single card. The coding and punching of each card was arranged to cover completely the information entered on the card.

The coding scheme is based on the following general headings:

- A. General subject
- B. Compounds described
- C. Author
- D. Date of publication
- E. Card serial number

A. General Subject Code. Space for direct coding of 32 general subjects is provided by 16 double holes along the upper edge of the card; shallow and deep punching are used. By using the intermediate punch, also, it would be possible to code a total of 48 subjects.

Numbers 1-16 are indicated by shallow punching, and numbers 17-32 by deep punching. The subject code was arranged so as to minimize overlapping of shallow and deep punching. Inconvenience due to overlapping has been negligible.

Table 4-1 gives some typical entries in the subject code. Since the subjects are direct coded, it is possible to code on each card all the subjects covered in the reference listed on that card.

B. Code for Compounds Described. This code is based on the various types of bonds and structural units found in organosilicon compounds. The code has been set up as a decimal classification. Nine important func-

TABLE 4-1. SUBJECT CODE

- 1. Analytical method (a new method)
- 2. Bibliography
- 3. Book
- .
- .
- .
- 10. Patent
- 11. Nomenclature discussion
- .
- .
- .
- 14. Resins
- 15. Review articles
- .
- .
- .
- 17. Structure investigation
- .
- .

TABLE 4-2. CHEMICAL CODE

1.00	Si-H
2.00	Si-Halogen (general)
2.10	Si-F
2.20	Si-Cl
	etc.
3.00	C-Halogen
4.00	Si-OH
5.00	Si-C
5.10	Si-Alkyl
5.11	Si-CH ₃
5.12	Si-C ₂ H ₅
5.20	Si-Aryl
5.21	Si-C ₆ H ₅
	etc.
6.00	Si-Si
7.00	Si-O-Si
8.00	Si-O-C
9.00	Si-Miscellaneous

tional or characteristic structural features were chosen as the main headings in the code. These nine headings were then further divided and subdivided. The general procedure is illustrated in Table 4-2.

A chemical compound is coded on a card by punching the appropriate classification numbers in the outer row of holes along the bottom of the card, in the section marked "Compound Index". For example, to code the class of compounds, Si-C₆H₅, its number, 5.21, is punched as 5 in the units field, 2 in the tenths field, and 1 in the hundredths field.

If more than one compound is mentioned in the reference, each such compound is coded in this same field. All cards bearing information about a given compound or group of compounds may then be mechanically selected from the file by procedures described in Chapters 2 and 3. Occasionally such a sort selects from the file not only all the cards desired but some extra ones, due to the superimposed coding. It has been found in actual practice that such extra cards are relatively few in number and are readily eliminated by inspection.

A mathematical analysis of superimposed coding is given in Chapter 21.

C. Author Code. If there is only one author, the first three letters of his surname are coded. If there are two or more authors, the first letters of the surnames of the first two and last authors are coded. The hole marked II is also punched to indicate plural authorship.

The holes assigned to coding author names are along the left edge of the card. The author code is based on the OIECB mnemonic code described by Casey, Bailey, and Cox². By applying the serial sorting procedure, de-

² Casey, R. S., C. F. Bailey, and G. J. Cox, *J. Chem. Ed.*, **23**, 495-9 (1946).

scribed in Chapter 2, the entire file may be arranged alphabetically by authors in a very short time. Some hand-sorting is required with multiple author cards.

Good results have also been obtained when this method of coding was used for locating a multi-author reference, the known-author's name of which was not that of the first author. In an actual case, it was desired to locate a reference known to bear a certain well-known author's name. After about ten-minutes searching the desired reference was located, and it was found that the known-author's name was the second among several other authors.

D. Date of Publication. The year of publication is coded in the holes located along the upper edge of the card and at the right of those used for the General Subject Code. The century of publication is directly coded in holes marked 17, 18, 19, while the decade and year are numerically coded using the 7-4-2-1 scheme described in Chapter 2.

E. Card Serial Number. Each card is assigned a serial number. This number is written on the card in the space marked I, and coded in the fields at the left edge of the card, as shown in Figure 4-1. Such a numbering system is extremely useful. The serial number may be used to identify the reference, as discussed in more detail below. Also, the last serial number assigned indicates the total number of cards in the file.

Use of the Punched Card File in Writing a Review Paper

The most important use of the file in writing the review paper was in selecting information pertinent to the subject matter discussed under the

The diagram illustrates a punched card layout for information on organosilicon compounds. The card is rectangular with various fields and indices marked. At the top left is a field labeled 'I'. Below it is a large central area. To the left of this area is a vertical strip labeled 'CARD NUMBER INDEX'. To the right is a vertical strip labeled 'AUTHOR INDEX'. At the top of the central area are two horizontal strips: 'SUBJECT INDEX' on the left and 'DATE INDEX' on the right. At the bottom of the central area is a horizontal strip labeled 'COMPOUND INDEX'. The card is filled with numerous small circles representing punch holes, some of which are numbered or labeled with letters like 'I', 'H', 'U', 'C', 'E', 'I', 'N', 'T', 'S', 'H', 'U', 'N', 'D', 'R', 'E', 'D', 'T', 'H', 'S', 'U', 'N', 'I', 'T', 'S'.

FIGURE 4-1. Card used for information on organosilicon compounds.

various headings and subheadings of the outline:

- I Introduction
- II. Nomenclature
- III. Methods for the synthesis of organosilicon compounds
 - A. Synthesis of organosilane and organochlorosilanes
 - B. Synthesis of organofluorosilanes
 - 1. By the Grignard reaction
 - 2. From silicones
 - 3. From organochlorosilanes
 - C. Synthesis of organochlorofluorosilanes
- IV. Behavior of classes of organosilicon compounds
 - A. Normal alkyls of the type SiR_4 , R_3SiSiR_3 , etc.
 - B. The alkylsilanes $\text{R}_n\text{SiH}_{4-n}$
 - C. Organosilanes with substituted alkyl and aryl groups
 - D. Alkyl and aryl halogenesilanes, $\text{R}_n\text{SiX}_{4-n}$
 - 1. Properties of organochlorosilanes
 - 2. Properties of organofluorochlorosilanes
 - E. The alkylalkyloxy- and aroxy-silanes $\text{R}_n\text{Si}(\text{OR})_{4-n}$
 - F. Organosilanols $\text{R}_n\text{Si}(\text{OH})_{4-n}$, etc.
 - G. Esters, $\text{CH}_3\text{COOSiR}_3$ and $(\text{R}_3\text{SiO})_2\text{SO}_2$
 - H. The linear and cyclic organosiloxanes
 - I. Silazanes and related groups
- V. The silicone polymers
- VI. Water-Repellent films
- VII. Special investigations of physical properties
- VIII. Isomerism
- IX. Physiological properties
 - X. Analytical methods
- XI. References

The cards bearing references to the desired subjects or types of compounds were selected from the punched-card file by sorting at the appropriate positions (holes). For example, before the section on Nomenclature was written, the cards were sorted in position number 11 in the subject index. In this way all the references pertaining to this topic were dropped from the file. It was then quite simple to draft this section directly from the cards. Likewise, the other sections of the publication were drafted by sorting first for the subject under consideration and subsequently rearranging the selected cards, by sorting procedures described in Chapter 2, into chronological order or according to subtopic to fit into the outline that had been previously prepared. For example, under Section V, entitled, "The Silicone Polymers", various physical properties of this class of compounds are dealt with. It was possible to separate first the silicone oil and rubber and resin topics from the file, and then arrange these so that the topics pertaining to viscosity, molecular weight, oil properties, resin properties, and rubber properties could be dealt with in their logical order.

The review paper also included a number of tables giving physical data

for various types of organosilicon compounds. In preparing the tables of organosilicon compounds the cards were sorted, using the Compound Index, to select the particular class of compounds desired. Each compound was then transcribed from the card together with its physical properties and card number. As a typical example, one of the tables contains the alkoxy- and aroxyorganosilanes. It was possible by sorting for 8 in the Compound Index to separate from the file those cards that had data pertaining to compounds containing the ether group. Then, by hand-sorting, these were arranged in the order desired.

Preparing the Bibliography

The punched-card file proved very helpful in preparing the bibliography of the review paper. During the preparation of the manuscript the card serial numbers were used as provisional reference numbers. When the final draft of the manuscript had been completed, the cards which had been used in its preparation were removed from the file and arranged alphabetically by author by serial-sorting the author index. The bibliography was then prepared directly from the cards by typing the references in alphabetical order by author. These references were numbered in sequence, and each such number was penciled on the corresponding card. To facilitate placing these final bibliography numbers in their proper places in the manuscript the cards were then arranged in order according to card serial number by sorting the card number index. It was then easy to scan the manuscript, find the card corresponding to each serial number used as temporary reference number, and replace the latter with the final bibliography number penciled on the card.

Advantages of Using Punched-Card Files for Preparing Papers, Reports, etc.

The application of punched-card techniques permitted the use of rapid mechanical methods to avoid tedious and time consuming hand-sorting and individual scanning of papers and cards.

The file on organosilicon compounds has also proved very useful in seeking and establishing correlations, such as cause-and-effect relationships. This is a matter of very great importance in preparing reports and papers concerned with previously unpublished data.

While discussing the advantages of punched cards for filing scientific and technical information, it should not be forgotten that the same file used in writing final reports, papers, etc., can also be used to good advantage in planning and executing experimental work. The author's file on organosilicon compounds was often consulted by himself and colleagues at the General Electric Co. when experimental work in that field was planned.

Briefer papers than the one referred to above have also been prepared using the same file. Thus, for a special report it was required that all the data pertaining to organosilicon compounds that contained the Si—OH group be obtained. The accompanying bibliography was to be placed in chronological sequence. The punched-card file was first sorted in the Compound Code field for 4 the code number for Si—OH. The cards isolated by this sorting operation were then arranged in chronological order by serial sorting the date of publication index (upper right field in Figure 4-1). The total time required for manipulating the punched cards was about 15 minutes.

In addition to entering numerical data and similar reference information on punched cards it is also possible to attach photographs, drawings, graphs, charts, microfilms, etc. directly to the cards. By sorting for a particular item or topic all corresponding photographs, drawings, and charts may be obtained in addition to the pertinent references and data. This is often helpful in writing papers as it may prove desirable to include photographs etc., which otherwise might be overlooked.

Keeping track of abstract references when setting up a file is facilitated by the following procedure: When searching abstract and journal indexes for a given compound or subject it is helpful to establish a special card—termed “Abstract Card”—on which all abstract references are listed together with any index information concerning the abstract. The compound or subject is edge-punched in the card in the usual way. When an original paper is read and a card prepared for that reference, the corresponding abstract notation on the Abstract Card is crossed with a light pencil mark. In this way the Abstract Card serves as a record of the search performed on a given subject or compound. A special hole is reserved to designate the Abstract Card and enables one to separate the Abstract Cards from the other cards in the file.

The hole designated by III in Figure 4-1 is used in the file just described to designate a company report.

Chapter 5

AN INTERNATIONAL CLASSIFICATION AND PUNCHED-CARD FILING SYSTEM FOR METALLURGICAL LITERATURE

MARJORIE R. HYSLOP

*Editor of A. S. M. Review of Metal Literature, American Society for Metals
Cleveland, Ohio*

AND

ALVINA WASSENBERG

*Research Librarian, Division of Metallurgical Research
Kaiser Aluminum & Chemical Corporation, Spokane, Washington*

Literature searching has long been accepted as a preliminary and highly important step in any research problem. It is tedious and time-consuming work. The excellent facilities offered by the various public, private and technical libraries to smooth the way of the researcher cannot be minimized. Nevertheless, a large part of the burden still remains the responsibility of the individual investigator.

Two problems face the technical researcher or investigator who must make up a file of technical information. First is a logical and usable analysis of the subject matter in his particular field of interest, either in the form of a list of subject headings or a classification outline. Second is the method of handling his files of information or literature references and the tools used therefor.

Recognition of this problem in the field of metallurgy was evidenced as early as 1947 by Guy and Geisler.¹ Response to the publication of this article was so enthusiastic that the eventual result was the formation of a joint committee by the American Society for Metals and the Special Libraries Association, whose function was to design a standardized classification system for metallurgical literature and a punched card filing system for use with it. The committee work was completed early in 1950 with the publication of the "ASM-SLA Metallurgical Literature Classification,"² ³

¹ Guy, A. G., and Geisler, A. H., "A Punch Card Filing System for Metallurgical Literature," *Metal Progr.*, **52**, 993 (Dec. 1947).

² ASM-SLA Metallurgical Literature Classification, prepared by a Joint Committee of the American Society for Metals and Special Libraries Association; American Society for Metals, Cleveland, Ohio, 1950.

³ Geisler, A. H., "How to Find Detailed Information When You Want It," *Metal Progr.*, **57**, 613 (May 1950).

and the underwriting of the manufacturing charges for a large supply of punched cards especially printed to accommodate the provisions of the classification.

In the intervening seven years the classification has been widely adopted, and the first printing of 2000 copies was exhausted early in 1956. It has been found during this time that the classification lends itself very well to organization of literature files that are not necessarily based on punched cards. To facilitate the organization of such files many abstracting services and technical journals, both in this country and abroad, print the classification code symbols in conjunction with titles of articles or abstracts. In this country all of the articles published in *Metal Progress*, *Transactions of the American Society for Metals* and all of the abstracts in the *A.S.M. Review of Metal Literature* are so coded.

In Italy, the Associazione Italiana di Metallurgia (Italian Association of Metallurgy), the Istituto Siderurgico Finsider (ferrous metallurgy) and the Istituto Sperimentale dei Metalli Leggeri (light metals) have adopted the ASM-SLA system.^{4, 5} These three organizations code their abstracting services and also the principal articles in their official journals with the ASM-SLA symbols, similar to the practice of a number of European publishers who print the symbols of the Universal Decimal Classification (U.D.C.).

In England, the classification and punched card system have been adopted by the Information Section of the British Iron and Steel Research Association,⁶ and in Germany by the Institut für Härterei-Technik, which also codes the principal articles and the abstracts published in its journal *Härterei-Technik und Wärmebehandlung*.

A new joint committee of the American Society for Metals and the Special Libraries Association was organized in 1955 to revise the classification and bring it up to date for a second printing (See Appendix A).⁷ The purpose of this revision is to provide for new fields of scientific knowledge (such as metallurgical aspects of atomic energy), as well as to expand some of

⁴ Classificazione Bibliografica Internazionale della Metallurgia, First Italian Edition of the ASM-SLA Classification; Italian Association of Metallurgy, Milan, Italy, 1955.

⁵ Documentazione Tecnica a cura del Centro Documentazione A.I.M.; *La Metallurgia Italiana*, 47, No. 6, 186 (June 1955).

⁶ Colinese, P. E., Why B.I.S.R.A. Has Adopted the American Society for Metals—Special Libraries Association (ASM-SLA) Metallurgical Literature Classification; *Aslib Proceedings*, 5, No. 4, 345 (Nov. 1953); *Metals Review*, 27, No. 12, 26 (Dec. 1953).

⁷ Committee Formed to Study ASM-SLA Literature Revision; *Metals Review*, Vol. 28, No. 7, July 1955, p. 14; 30, No. 26 (Feb. 1957); see also L. S. Foster, "Revision of the ASM-SLA Code for Metallurgical Literature." Paper presented before the Division of Chemical Literature, 131st National Meeting of the American Chemical Society, Miami, Florida, April 8, 1957.

the sections in somewhat greater detail. Additions and revisions are made in such a manner that the general outline of the first edition can be retained so that existing files are not invalidated. A parallel committee of the Italian Association of Metallurgy is working closely with the American committee in this revision so that the new edition of the classification will represent a joint undertaking which can eventually be adopted as an international standard for metallurgical literature classification.

While there are no limitations on the size of literature files that can be maintained using the classification code symbols, the limitations for a punched card file are somewhat narrower. Sales of the punched card, which was especially designed for use with the system, have continued at a steady rate; by early 1956 sales had reached a total of more than two million cards. While it is difficult to estimate the number of individual files represented by this total, a conservative guess would be in the neighborhood of several hundred at least.

The upper limit for hand sorting marginal punched card files is estimated to be about 10,000 literature references before the system breaks down because of the limitations of hand sorting and needling. The punched card system is therefore primarily suited to the needs of the individual researcher whose files are more likely to number somewhere between five hundred and a couple of thousand references. It is also proving useful to the technical librarian whose interests are largely confined to some aspect of metallurgical science and technology, and an installation in a company library will be described in some detail later in this chapter.⁸ ⁹ A system maintained by a plant metallurgist was described before the Metals Division of the Special Libraries Association by E. C. Wallace in October 1955.¹⁰

The Metallurgical Literature Classification

Previous to the work of the ASM-SLA committee, no thorough analysis of the subject of metallurgy as an entity was available. Existing classification systems, such as the Dewey Decimal, Universal Decimal, and various library cataloging systems, were designed to accommodate all fields of science and technology, with the result that metallurgical interests were scattered somewhat promiscuously among a variety of other subjects. What attempts had been made to treat metallurgy as a separate science were either inadequate or impractical.

No one method of breaking down the subject of metallurgy can be uni-

⁸ Wassenberg, Alvina, Experience With the ASM-SLA Classification of Metallurgical Literature. Paper presented before Metals Section, S-T Division, Special Libraries Association, Detroit, Mich., October 1951.

⁹ Edelman, David L., "Library Use of New Indexing System," *Metal Progr.*, 59, 526 (April 1951).

¹⁰ Wallace, E. C., "Use of the ASM-SLA Metallurgical Literature Classification System by Industrial Metallurgists," *Special Libraries*, 47, No. 3, 114 (March 1956).

versally useful to every metallurgist and every librarian. One person may be primarily interested in metallurgical processes, another in metallic properties, a third in specific metals and materials, a fourth in a specific metal-product form (such as tubing, sheet, wire), and still a fifth may wish to classify his data by equipment used.

With this in mind, the committee decided to provide three parallel and independent classification outlines or indexes, one a "Processes and Properties Index," one a "Materials Index," and the third a so-called "Common Variables Index." A condensed outline of the entire classification is given in Appendix A, where changes and additions appearing in the second edition are indicated by italics. In addition, the punched card provides for an Author Index and an optional Date Index.

In preparing these indexes, the committee decided to provide a systematic, logical and practical breakdown by subject without attempting to fit the various headings into any preconceived coding system, such as a decimal arrangement. The fundamentals of the punched card, however, were borne in mind so that the various headings and subheadings could be accommodated on the card with an appropriate coding system consisting of combinations of letters and numbers and semantic symbols. It will not be necessary to reproduce the complete classification here, but enough will be given to illustrate its general outline and arrangement.

The Processes and Properties Index resolved itself neatly into twenty main divisions (termed for convenience "first-order divisions") as follows:

- A — General Metallurgical
- B — Raw Materials and Ore Preparation
- C — Nonferrous Extraction and Refining
- D — Ferrous Reduction and Refining
- E — Foundry
- F — Primary Mechanical Working
- G — Secondary Mechanical Working
- H — Powder Metallurgy
- J — Heat Treatment
- K — Joining
- L — Cleaning, Coating and Finishing
- M — Metallography, Constitution and Primary Structures
- N — Transformations and Resulting Structures
- P — Physical Properties and Test Methods
- Q — Mechanical Properties and Tests Methods; Deformation
- R — Corrosion
- S — Inspection and Control
- T — Applications of Metals in Equipment
- U — Allied Fields
- V — Materials (Subdivided in Materials Index)

TABLE 5-1. SAMPLE OF SECOND AND THIRD-ORDER SUBDIVISIONS IN PROCESSES AND PROPERTIES INDEX

A — GENERAL METALLURGICAL		C — NONFERROUS EXTRACTION AND REFINING	
2. History	d. Screening	21. Smelting	
3. Education	e. Wet sizing (classification)	a. Blast furnace	
4. Statistics and economics	f. Air classification	b. Converting (to include bessemer)	
5. Plant practice	14. Concentration and beneficiation	c. Reverberatory processes	
a. Materials handling	g. Gravity	d. Electric furnace	
6. Industrial relations	h. Flotation	22. Distillation	
7. Health and safety	j. Magnetic	g. Reduction	
8. Secondary metals, scrap and waste disposal	k. Chemical (leaching, etc.)	h. Refining	
9. Research organizations	m. Agglomeration tabling	23. Electrolytic processes	
10. Glossaries, definitions, trade names, directories	n. Electrostatic	n. Electrowinning	
B — RAW MATERIALS AND ORE PREPARATION	p. Settling, thickening, filtration	p. Electrorefining	
	15. Roasting and calcining	24. Cyanidation	
	16. Sintering and nodulizing	25. Vacuum refining	
10. Ore deposits and raw materials reserves	17. Briquetting	26. Reduction by metals (Thermit processes, etc.)	
11. Sampling (For Assaying, see S11a)	18. Fuels technology*	27. Cementation (copper on iron, etc.)	
12. Mining	n. Solid	28. Separation of metals	
n. Open pit	p. Liquid	g. Parting	
p. Underground	q. Gaseous	h. Liquation	
q. Hydrauliclicking	19. Refractories technology*	29. Amalgamation	
13. Crushing, grinding and sizing	21. Fluxes and slags	1. Carbonyl reduction	
a. Primary crushing	22. Addition agents	2. Hydride decomposition	
b. Secondary crushing and grinding	n. Ferro-alloys	4. Halide decomposition	
c. Milling	p. Other metal additions	5. Ingot casting	
	q. Ores	n. Tapping	
	r. Scrap	p. Teeming	
	s. Nonmetallics (coke, carbon, etc.)	q. Continuous casting	
	a. Oxygen		

* Limited to fuels and refractories in general. Use of fuels and refractories in a specific metallurgical process (melting, heat treating, welding, etc.) should be indexed in the appropriate process section and cross-indexed to fuels or refractories in the Common Variables Index.

Each of these main divisions was further broken down into some ten to twenty second-order divisions. When complexity of the subject matter required, these second order divisions were further subdivided into third orders. An example of the breakdown into second and third orders is given for the first three principal divisions of the classification, namely, A — General, B — Raw Materials and Ore Preparation, and C — Nonferrous Extraction and Refining. (See Table 5-1). Provision was also made on the punched card for fourth orders to take care of the detailed breakdown required by specialists in certain fields; none of these fourth-order divisions, however, were provided by the committee. Ample provision was made for expansion of the outline in any direction.

Formulation of the Materials Index presented some special problems. While materials in this metallurgical index are confined to metals, the metallurgist has various and sometimes contradictory or overlapping ways of grouping them. Metallic alloys may be grouped by base metal (aluminum

alloys), by composition (aluminum-copper-magnesium alloys), or by specific properties or uses (heat-resisting alloys, magnetic materials, bearing metals). The ferrous alloys (steels) present still further difficulties since references may be required to such diverse and overlapping groups as openhearth, bessemer or crucible steel, carbon steels, alloy steels, tool steels, stainless steels, and cast iron. Provision for indexing by either composition or group or both was made, as will be explained later in the section on the punched card.

The Common Variables Index is a miscellaneous collection of factors which modify either the Processes and Properties Index or the Materials Index, or which refer to the physical characteristics of the publication. Such subject headings as equipment, theory, high and low temperature, are common to a large proportion of the first- and second- and even third-order subdivisions in the Processes and Properties Index. Such "metal forms" as castings, forgings, tubing and coated metals might be considered as modifying the Materials Index. Physical characteristics of the publication include type of literature (theory, research, review, plant description), form (patent, specification, report, book) and language.

The Common Variables Index is one of the distinguishing features contributing to the versatility of the punched-card system. In conventional card files, its subdivisions can be incorporated as additional subheadings in the Processes and Properties and the Materials Indexes, or carried as separate main headings.

For purposes of coding published articles or abstracts, the symbols used for each of the three indexes are sufficiently distinctive that they need be separated only by commas; if a more pronounced division of the three indexes is desired, semicolons or colons may be used to separate the series of symbols required for each index.

One disadvantage of the conventional alphabetical subject index file is that many duplicate cards must be made, each one carrying a single entry. For example, a reference may be concerned with furnaces for heat treating aluminum alloy castings. Indexing entries would be required for "furnaces," "heat treating," "aluminum," and "castings." In the punched card system all of these entries can be indexed on one card, and all may be considered as primary headings, and not subordinated as subheadings.

Design of the Punched Card

The card selected for the ASM-SLA system is a standard 5 x 8-inch card manufactured by E-Z Sort Company and marketed by Lee F. Kollie Associates, Inc. of Chicago. (Figure 5-1.) A double row of holes is punched on all four edges of the card, with a row of six double holes to the inch instead of the more usual four per inch.

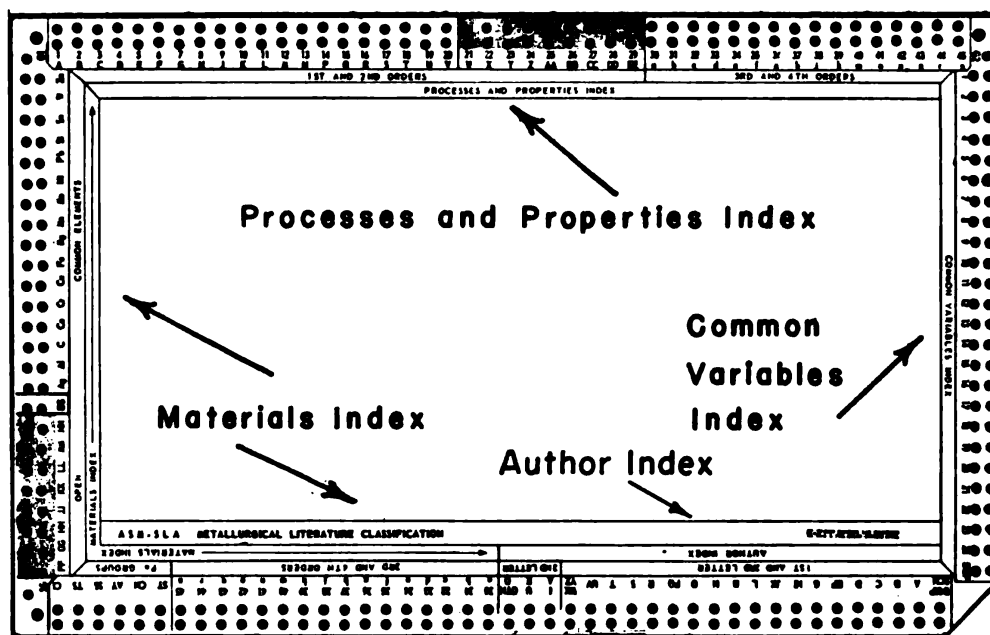


Figure 5-1. Punched card specially designed for use with the ASM-SLA classification of metallurgical literature, showing the location of the four principal indexes.

The Processes and Properties Index occupies the top edge of the card, the Materials Index the left side and bottom left edge, the Common Variables Index the right edge, and the Author Index the bottom right edge. If desired, a fifth index by date can be incorporated in a portion of the space allotted to the Common Variables Index. Both direct indexing (a single hole assigned to a single specific criterion) and indirect indexing (permutations of two or more holes to designate a single criterion) are used as dictated by the exigencies of the four indexes.

Processes and Properties Index. The main divisions of this index are directly coded by the alphabetical designations A through V. Additional divisions coded W, X, Y, Z, AA, BB, CC, DD, and EE are open for future expansion. These letters are assigned to the lower or deep row of holes across the top of the card, starting at the left. Each of these "first-order" divisions carries a specific meaning: A — General Metallurgical, B — Raw Materials and Ore Preparation, etc. Considered alone, it constitutes a direct index.

Each of these main headings is followed by a series of second-order headings designated by the numerals 1 through 29, assigned to the upper or shallow row of holes at the top of the card. Since shallow hole No. 2 in Division A (history) has a meaning different from shallow hole No. 2 in Division C (hydride decomposition), the combination now constitutes an indirect index utilizing combinations of deep and shallow holes to indicate a single meaning. Figure 5-2 shows a card with an abstract mounted on it,

coded and slotted. The reference is coded B10 and A9 in the Processes and Properties Index; a deep notch in B and shallow notch in 10 stands for ore deposits and raw material reserves (mineral resources of Turkey) and a deep notch in A and shallow in 9 stands for research organizations (Government Research Institute).

Third and fourth orders are provided in the Processes and Properties Index to give a finer subdivision of the subject matter. This also is an indirect index, using the lower case letters "a" through "s" assigned to the deep holes at the top right of the card for third orders and the corresponding shallow holes designated 30 through 45 for fourth orders.*

The corner holes, marked G_1 and G_2 , are for "general" subdivisions of first- and second-order divisions, respectively. For example, a reference coded C- G_1 would indicate that it refers to the general subject of "Non-ferrous extraction and refining" rather than any specific divisions of this general subject.

In the indirect type of index, such as the Processes and Properties Index, the number of criteria that can be simultaneously coded is strictly limited, and it is recommended that no more than three first-order divisions be coded on any one card. (As many second-order divisions as may be desired can be indexed under one main division.) For this reason, Section V

* In arranging the classification, the numeral which occupies the same space as a letter designation is omitted from the second-order subdivisions under the first-order division, since it would be automatically notched out when the first-order division is coded. (See Table 5-1.) Thus in Division A, the numbering of the second orders starts with No. 2 instead of No. 1; No. 1 is automatically eliminated when the letter A is notched.

Also the ideal arrangement of this coding system would have been to start each second-order subdivision under a main division with No. 1, and each third order with the letter "a," but such an arrangement would have resulted in overcrowding in the first portion of the code. To avoid such overcrowding, the arrangement of code symbols in second, third, and fourth orders is staggered. Staggering in second-order subdivisions is roughly by multiples of ten—Division A starts with No. 2, Division B starts with 10, Division C with 21, and Division D jumps back to No. 1 again for its first designation. After No. 29 is reached in any one main division, the coding drops back to No. 1; for example, in Division C, second-order subdivision No. 29 "amalgamation" is followed by a parallel second-order subdivision No. 1 "carbonyl reduction."

Staggering in third orders is in multiples of five letters. In Division C, for example, the first third-order subdivision under the first second order "21 Smelting" is coded with the letter "a" ("a. Blast furnace"). The first third-order under "22. Distillation" is "g. Reduction" and the first third-order under "23. Electrolytic processes" is "n. Electrowinning." Had third-order subdivisions been provided under "24. Cyanidation," they would have started again with the letter "a". After the letter "s" is reached, the coding goes back again to the beginning of the alphabet. (See "B22s. Nonmetallic addition agents," followed by "B22a. Oxygen as an addition agent.")

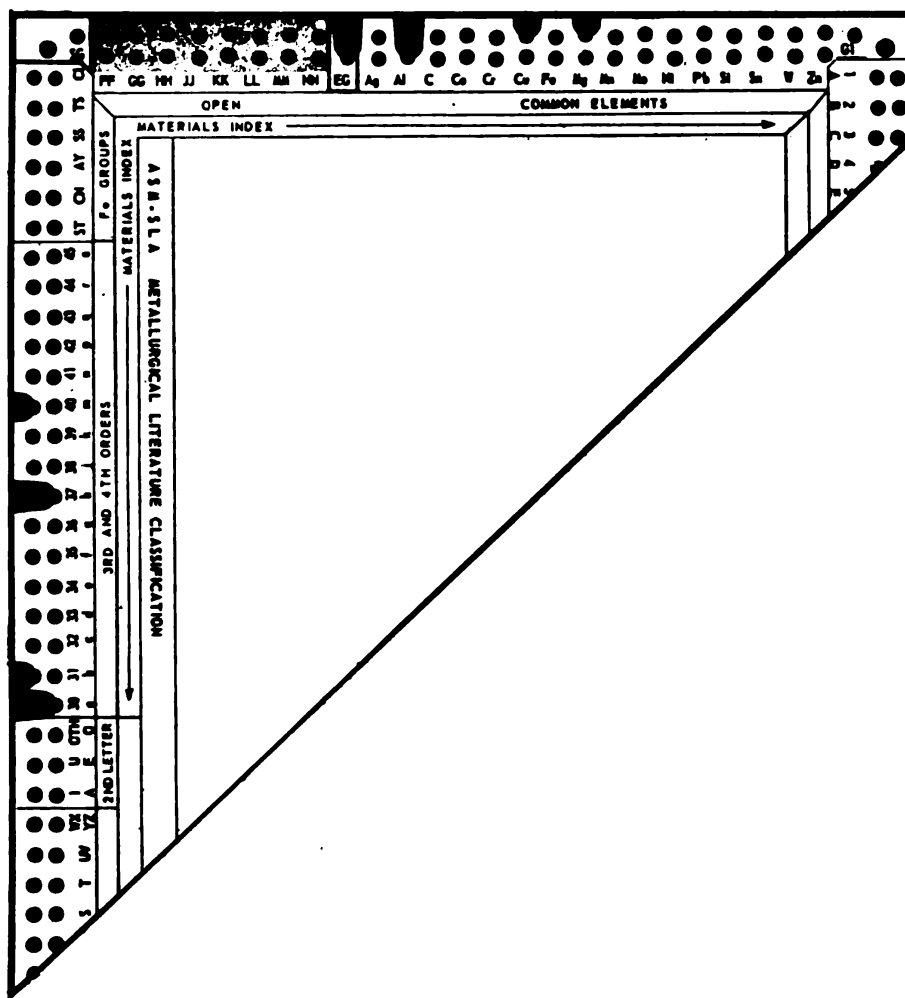


Figure 5-3. Three examples of Materials Index Coding. Aluminum-copper-magnesium alloys indicated either by a combination of a deep hole in Al and shallow in Cu and Mg, or by Al-h-40 (duralumin), utilizing the third and fourth orders section. Antimony, not considered a common element, is coded EG-a-31.

Index (letters "a" through "s" and numbers 30 through 45). A completely coded outline for such subdivisions of the alloys of the common metals is provided in the classification outline. For example, the aluminum alloy known as duralumin can be coded Al-h-40, according to the classification outline. In Figure 5-3 this is notched deep in Al in the Common Elements Section, and deep in h, and shallow in 40 in the third- and fourth-order section.

This third and fourth order section is also utilized for indirect indexing of the less common elements which are not coded by chemical symbols. It will be noted that a double hole designated "EG" appears just below the center of the left edge. This designation (for "Elements Grouped") includes

all the other elements of the periodic table, reached by combinations of the lower case letters and numerals in the third- and fourth-order section. For example, in Figure 5-3, the element "antimony" is coded by notching EG-a (deep) and 31 (shallow).

The various types of steels and irons are provided for in the section at the bottom left in the card designated "Ferrous Groups." The divisions are as follows: ST—Steels, CN—Carbon Steels, AY—Alloy Steels, SS—Stainless Steels, TS—Tool Steels, CI—Cast Iron and Cast Steels. This again constitutes a direct index. (See Figure 5-2, notched in Fe for iron and ST for steels generally.)

For indexing materials by special properties or application, the lower left corner hole designated "SG" is provided, and is used in conjunction with the third order lower case letters. For example, heat-resisting alloys are coded by notching SG-h; bearing metals SG-c, etc. In these "Special Groups" fourth orders are open for further subdivision as required by the user.

Common Variables Index. As stated above, the Common Variables Index makes full use of the unique faculty of a punched-card system to record several concepts simultaneously and thus provide selection of information by an interrelation of ideas. It is a miscellaneous collection of factors which modify a number of separate headings in the other indexes.

The Common Variables Index is accommodated on the right edge of the card and may be handled in various ways at the option of the user. Three methods, using either the direct or indirect coding principle, are suggested in the book containing the complete classification and explanation published by the American Society for Metals.

Author Index. A new method of indexing authors was developed which represents an ideal combination of simplicity, ease of operation, selectivity, and economy of space. Studies were made of the frequency of occurrence of the various letters as the first, second, and third in metallurgists' surnames, using the author lists from the ASM Review of Metal Literature and the British Metallurgical Abstracts. Results showed that the heaviest concentration of specific letters occurs in the second letter; in fact, the vowels A, E, I, O, U, plus the consonants H, L and R constitute about 90 per cent of the second letters. It was therefore concluded that, for greatest selectivity, it is even more important to index the third letter of the author's name than the second.

The Author Index is accommodated on the bottom right edge of the card (see Figure 5-4). The first letter of an author's surname is indexed in the deep holes in the larger field designated "First and Third Letter." The third letter is indexed in the shallow holes in this field. The second letter is indexed by a deep or shallow notch as required on the small field designated

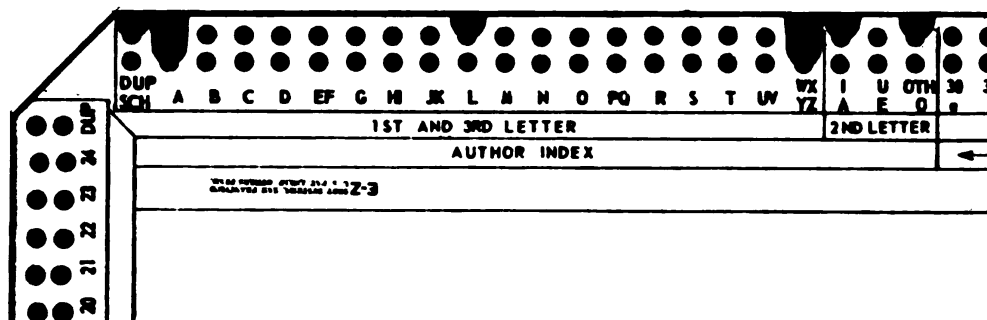


Figure 5-4. Indexing of two authors on one card. Wilson, coded W, I, L, and Adams, coded A, D and DUP, for duplication of the first and third letter.

“Second Letter.” In this second-letter field the vowels are each indexed separately, while all of the consonants are indexed by notching the shallow hole marked “Other.”

The common combination “Sch” is considered as a unit and indexed separately. The following fourth letter of the name is then indexed in the second-letter field, and the fifth letter is indexed in the proper shallow hole in the first-and-third letter field.

The shallow hole marked “Dup” (corresponding to the deep hole “Sch”) is provided for instances when the first and third letter are the same.

Capacity for Expansion. It will be noted that there are two shaded areas on the card (Figure 5-1), one in the Processes and Properties Index covering the code letters W through EE, and the second in the Materials Index, letters FF through NN. These are provided for addition of any subject as required by the future growth of science and technology. They may be used for the addition of material in fields other than metallurgy, or for advancing specialized fields to more important first-order positions. For example, a metallurgist who has a large file of references on electrolytic refining may wish to advance C-23 to create W — Electrolytic Refining, or one who is particularly interested in materials handling may save needling time and provide for more detailed subject breakdown by assigning first-order division X to materials handling instead of notching and needling through A-5-a.

The potential capacity of the classification system is indicated by the following figures for the Processes and Properties Index alone: 29 first-order divisions; 27 subdivisions under each of these, making a total of 756 second orders; 16 subdivisions under each of these, or a total of 12,096 third orders; and 15 subdivisions under each of these, or a total of 181,440 fourth orders.

Some of the open areas will be used by the Committee now engaged in revising the classification. For example, in the five years intervening since the first publication, experience has shown that certain nonmetallic ma-

materials used in metallurgical processing should have some place in the outline, and a new section in the Materials Index will be opened for this category. Many new subsections will also be opened under the various main divisions of the Processes and Properties Index and the Common Variables Index. A good example of the capacity for expansion of the system is the fact that all of the metallurgical aspects of nuclear sciences (which were unforeseen five years ago but now constitute a vast and widely assorted segment of the literature) will be accommodated in the revision merely by adding appropriate subdivisions to the various sections of all three indexes.

The Workbook. Since one of the most useful features of the ASM-SLA classification system is its capacity for expansion, a highly adaptable method of working with the classification outline is demanded. To make most efficient use of the classification, therefore, a looseleaf "workbook" is provided. In this book each of the main divisions of the classification is printed on looseleaf sheets that can be thumb-indexed for ready reference. Existing second- and third-order subdivisions are also printed, together with a complete list of coding symbols for all of the open second, third and fourth orders, thus providing space for additions to the classification at the will of the user. A sample page from this book is reproduced as Figure 5-5.

Experience in a Company Library

In early 1950, a new Research Library was being organized at Kaiser Aluminum and Chemical Corporation's Division of Metallurgical Research in Spokane. At the same time the American Society for Metals and the Special Libraries Association were jointly preparing the punched-card system for coding metallurgical literature described in the first part of this Chapter. The need for an abstracting system in the Kaiser Research Library, coupled with the Librarian's previous experience with punched cards, made this ASM-SLA system seem desirable. Furthermore, this system could be readily adapted to a young organization such as the Kaiser Research group. These factors guided the decision to install the system as soon as it was completed.^{8, 9}

Word reached Spokane that the ASM-SLA Classification would be available at the time of the Special Libraries Association Convention in Atlantic City in June, 1950. Of the two copies of this Classification of Metallurgical Literature which were available there, one went to the Northwest for study and adaptation in the Kaiser Laboratories.

In mid-July, 1950, the various items needed to start the punched card system began to arrive, and abstracting was begun on all literature in the Library, except the books and periodicals. In about two months there was a fairly large file of technical information recorded on classification cards. Then the work of actual coding began and many hours were spent going over the cards, tentatively coding them, and then exchanging ideas between

Third-Order Divisions

B — Raw Materials and Ore Preparation

- | | |
|------------------------------|-------------------------------------|
| 10. Ore deposits & mat. res. | 13. Crushing, grinding, and sizing |
| a. | a. Primary crushing |
| b. | b. Secondary crushing and grinding |
| c. | c. Milling |
| d. | d. Screening |
| e. | e. Wet sizing |
| f. | f. Air classification |
| g. | g. |
| h. | h. |
| j. | j. |
| k. | k. |
| m. | m. |
| n. | n. |
| p. | p. |
| q. | q. |
| r. | r. |
| s. | s. |
| 11. Sampling | 14. Concentration & beneficiation |
| g. | g. Gravity |
| h. | h. Flotation |
| j. | j. Magnetic |
| k. | k. Chemical |
| m. | m. Agglomeration tabling |
| n. | n. Electrostatic |
| p. | p. Settling, thickening, filtration |
| q. | q. |
| r. | r. |
| s. | s. |
| a. | a. |
| b. | b. |
| c. | c. |
| d. | d. |
| e. | e. |
| f. | f. |
| 12. Mining | 15. Roasting and calcining |
| n. Open pit | n. |
| p. Underground | p. |
| q. Hydraulicicking | q. |
| r. | r. |
| s. | s. |
| a. | a. |
| b. | b. |
| c. | c. |
| d. | d. |
| e. | e. |
| f. | f. |
| g. | g. |
| h. | h. |
| j. | j. |
| k. | k. |
| m. | m. |

Figure 5-5. A sample worksheet showing blanks for additional entries in third-order rank in Section B on Raw Materials and Ore Preparation.

library and technical staff members. There were many conflicting ideas on how to code, what to code, etc., but after checking about a thousand abstract cards, things began to smooth out and the system began to take shape in the Library.

In many organizations, where an abstract file is already in existence transferring the information to the new cards would be almost an impossibility due to the amount of clerical work involved. Microfilm strips containing the abstract could be inserted in the cards as a timesaver, but would necessitate special equipment, such as a reader, special cards, etc. In a newly formed organization, starting from scratch, the installation is somewhat simpler. Consequently, it was relatively easy in the Kaiser Laboratories. Many industries might find the ASM-SLA Classification easier to use than do those in the light metals field. In dealing with the aluminum literature it was found necessary to do a great deal of expanding into the worksheets provided with the ASM-SLA Classification books.

The ASM-SLA system was adopted for indexing all reports on research work conducted in the Division, as well as supplemental information contained in published articles related to similar projects. Thus the punched cards in the Kaiser Library now include references to papers from many sources such as the Bureau of Mines, U. S. Bureau of Standards, NACA Reports, PB Reports, etc.

No attempt has been made to code books and periodicals, because they are cataloged according to the Library of Congress classification. Of course, some technical articles in the periodicals are abstracted and incorporated in the ASM-SLA system.

As a safeguard against overloading the cards with extraneous material of little value, Kaiser research staff members help the Librarian to select published articles and abstracts for inclusion in the file. A number of the ASM-SLA Classification books have been assigned to department heads for use in coding reports and articles. The ASM-SLA "Worksheets" are also provided for each department head.

As new subjects are included in the system, the items are added to the list on the Worksheets. A supplementary index is issued when enough items have been submitted to justify the publication of a new list. Copies are issued to all members on the staff who use the ASM-SLA Metallurgical Classification.

Figure 5-6 shows a page in an abstract bulletin with code annotations in the margin and a card punched and abstracted accordingly. Only the "Processes and Properties" and "Author" indexes were needed for this particular reference. Of the code symbols jotted in the margin, Q24 at right stands for Plastic Deformation (24) in the Mechanical Properties section (Q). Q3 stands for Creep, M26s for Crystal Imperfections, and M22g for X-Ray Diffraction.

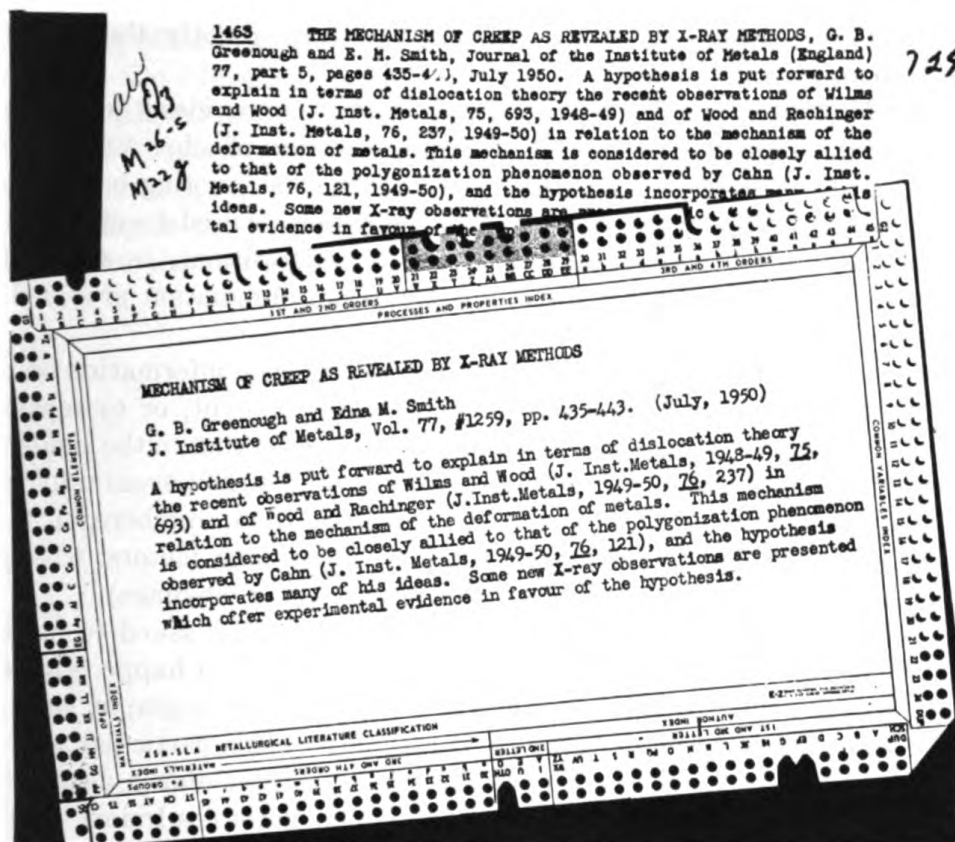


Figure 5-6. Page from an abstract bulletin with code notations in margin and punched card slotted accordingly.

Occasionally, an investigator finds that the standard classification does not adequately index a particular article or report. When this occurs, he discusses the matter with the head of his department, and the two of them select a new subheading to meet their needs, decide under which of the main classification headings it belongs, and submit their suggestions for expansion of the index to the Librarian for final approval. The new subheading is then entered in all copies of the Worksheets.

For example, several new classifications were needed in connection with remelt studies. In the aluminum industry, alloying and ingot casting are not considered part of the refining process as they are in some other non-ferrous industries. The nature of remelt practices indicated that they could be added best to the section of the "Processes and Properties" index designated "E—Foundry." Three new subheadings were entered in the Worksheets—namely, "E7. Ingot Casting," "E8. Fluxes and Fluxing," and "E9. Alloying."

Because of the uncertainty of fields and phases of research in which the Division will be engaged ten years hence, it was deemed advisable to minimize the use of fourth-order indexing and new headings. Fourth-order

divisions are added only when unavoidable; consequently there is still much room for expanding the system.

The "Materials Index" section of the classification provides a good breakdown for coding aluminum alloys. With the addition of alloy 50S (K-150) as "Al-g-34," this section is adequate to cover most existing commercial alloys. The "Common Elements Index," however, is used to classify experimental alloys according to their chemical compositions, as well as for other metals and alloys that have no commercial designation provided in the standard index.

The "Common Variables Index" may be used to code information on the construction and use of specialized laboratory equipment, or to separate reports on different phases of a research program according to the influence of various factors. Although the entire section will be used eventually, the divisions which are currently finding most use are those numbered 1, 2, 3, 4, 10, 11 and 12 (equipment and processes, influencing factors, wrought metal forms, type of literature, form of literature, and language).

If a literature search is made by a staff member, he is asked to furnish the Library with a copy of every abstract he prepares. If it happens to be a very extensive search, he is requested to compile a bibliography and an MR (Miscellaneous Report) number is assigned to this particular search. This bibliography is then coded for the subject matter and is also coded 11-15 to indicate "bibliography" in the Common Variables Index.

From a time-saving viewpoint, it is necessary to keep the cards in a readily accessible filing cabinet. Since no filing is required, this system can be used by technical men in their personal files. Of course, the coding and notching of the cards must be kept up-to-date or the system defeats itself. Furthermore, the cost of the cards themselves is too great for them to be used merely as regular abstract cards.

One trend which can cripple the entire system is the inclusion of too much material of marginal interest. If the people on the technical staff proceed to code all literature they may encounter, the Library will find itself swamped with journals of no probable lasting value waiting to be abstracted. However, an explanation of the requirements for a good abstract file will generally result in full cooperation of all participants. In order to build a good technical library, the wholehearted support of the staff is necessary. As Ralph H. Phelps of the Engineering Societies put it during a New York SLA Meeting, "know what you need and get it if available" with the corollary, "avoid getting and keeping what you do not need."

Another question that arises in connection with the system looks to the time when the number of cards will be too cumbersome to handle. There are several solutions to this problem as we see it now. One would be to weed out abstracts covered by the various abstracting services, such as Chemical

Abstracts, Metallurgical Abstracts, etc., when the annual indices are received in the Library. Another practice would be to subdivide cards by subjects as listed in the ASM-SLA Classification of Metallurgical Literature under the "Processes and Properties Index." This, of course, would mean needling through about 20 to 25 sets of grouped cards, but it might be time-saving in the long run. An efficient mechanical sorting device might well be the ultimate solution.

APPENDIX A

ASM-SLA CLASSIFICATION, INTERNATIONAL (SECOND EDITION), 1958

Condensed Outline of First and Second Orders

The following condensed outline of first and second orders is still tentative and some changes will appear in the final published version of the complete classification outline. Also it should be noted that the preceding de-

scriptive text is based on the first edition and cites a few examples which are no longer pertinent. The general principles, however remain unchanged. Changes from first edition are indicated by *italics*.

Processes and Properties Index

A — General Metallurgical	C — Extraction and Refining	E — Foundry
2. History	19. <i>Chemical extraction</i>	10. Melting techniques
3. Education	21. <i>Nonferrous smelting</i>	11. Sand casting
4. Statistics and economics	22. Distillation	12. Permanent mold casting
5. Plant management	23. Electrolytic processes	13. Die casting
6. Industrial relations	25. <i>Nonferrous vacuum methods</i>	14. Centrifugal casting
7. Health and safety	26. Reduction by metals <i>and gases</i>	15. Precision investment casting
8. <i>Industrial pollution and waste disposal</i>	27. <i>Precipitation (replacement)</i>	16. Other methods
9. Research and development	28. Separation of metals	17. Pattern making
11. <i>Conservation of resources</i>	29. Amalgamation	18. Sand technology
12. <i>Professional activities</i>	1. <i>Decomposition</i>	19. Molding
13. <i>Laws and legal actions</i>	5. <i>Nonferrous melting and casting</i>	21. Coremaking
14. <i>Literature techniques</i>	6. <i>Isotope separation</i>	22. Rigging
15. <i>Audio-visual techniques</i>		23. Pouring
B — Ore and Raw Material Preparation	D — Iron and Steelmaking	24. <i>Casting cleaning</i>
12. Mining methods	1. Blast furnace practice	25. Metallurgical control
13. Crushing, grinding and sizing	2. Openhearth practice	26. <i>Casting reclamation</i>
14. Concentration (beneficiation)	3. Bessemer process	
15. Roasting and calcining	5. Electric arc processes	F — Primary Mechanical Working
16. <i>Sintering processes</i>	6. Electric induction practice	21. Preparation for working
17. <i>Fuel distillation and gasification</i>	7. <i>Multistage processes</i>	22. Forging (hammer and press)
18. <i>Processing of fuels</i>	8. <i>Special processes</i>	23. Rolling
19. <i>Processing of refractories</i>	9. Ingot casting	24. Extrusion of rod, shapes and tube
23. <i>Processing of scrap</i>	10. <i>Oxygen furnace processes</i>	25. Swaging
24. <i>Water treatment</i>	11. <i>Physical chemistry of iron and steel-making</i>	26. Tubemaking (except extrusion)
25. <i>Production of industrial gases</i>		27. Drawing of rod and shapes

- 28. Wire drawing
- 29. Finishing operations
- G — Secondary Mechanical Working
 - 1. Presswork in general
 - 2. *Shearing*
 - 3. Stamping
 - 4. Drawing
 - 5. Impact and cold extrusion
 - 6. Bending
 - 9. Stretch forming
 - 10. Heading
 - 11. Roll forming
 - 12. Thread rolling
 - 13. Spinning
 - 14. Bulging
 - 16. Die sinking
 - 17. Machining, *mechanical*
 - 18. Grinding
 - 19. *Finish machining*
 - 22. Flame and arc cutting
 - 23. *Finish forming*
 - 24. *Special machining methods*
- H — Powder Metallurgy
 - 10. Powder production
 - 11. Powder properties
 - 12. Mixing and blending
 - 14. Molding and compacting
 - 15. Sintering
 - 16. Post-sintering operations
 - 17. *Fiber metallurgy*
- J — Heat Treatment
 - 21. Homogenization
 - 22. Austenitizing
 - 23. Annealing
 - 24. Normalizing
 - 25. Patenting
 - 26. *Quenching*
 - 27. Precipitation hardening
 - 28. Case hardening
 - 29. Tempering
 - 1. *Residual stress control*
 - 2. Heating and cooling methods
 - 3. Other carbon diffusion processes
 - 4. Surface alterations
 - 5. *Hardenability*
- K — Assembling and Joining
 - 1. Arc welding
 - 2. Gas welding
 - 3. Resistance welding
 - 4. Thermit welding
 - 5. Forge welding
 - 6. Other welding processes
 - 7. Soldering
 - 8. Brazing
 - 9. *Process control*
 - 11. Bonding metals to nonmetals
 - 12. Adhesive joining
 - 13. Mechanical *assembling*
- L — Cleaning, Coating and Finishing
 - 10. Mechanical cleaning and polishing
 - 12. Chemical cleaning and polishing
- M — Metallography, Constitution and Primary Structure
 - 20. *Specimen preparation*
 - 21. Microscopy
 - 22. Diffraction methods
 - 23. Other techniques
 - 24. Phase diagrams
 - 25. *Atomic and molecular* structures
 - 26. Crystal structures
 - 27. *Primary* microstructures
 - 28. Macrostructure
- N — Transformations and Resulting Structures
 - 1. Diffusion
 - 2. Nucleation
- 13. Electrochemical cleaning and polishing
- 14. Chemical *conversion* coating
- 15. Diffusion coating
- 16. Dip coating
- 17. Electroplating
- 18. Electroforming
- 19. *Anodic coating*
- 21. Cathodic oxide coating
- 22. Cladding or bonding
- 23. Metal spraying
- 24. *Coating by weld-deposition*
- 25. *Coating by vapor-deposition*
- 26. Painting and organic coating
- 27. Ceramic coating (vitreous enameling)
- 28. *Coating by chemical reduction*
- 29. Other methods

APPENDIX A

ASM-SLA CLASSIFICATION, INTERNATIONAL (SECOND EDITION), 1958

Condensed Outline of First and Second Orders

The following condensed outline of first and second orders is still tentative and some changes will appear in the final published version of the complete classification outline. Also it should be noted that the preceding de-

scriptive text is based on the first edition and cites a few examples which are no longer pertinent. The general principles, however remain unchanged. Changes from first edition are indicated by italics.

Processes and Properties Index

A — General Metallurgical	C — Extraction and Refining	E — Foundry
2. History	19. <i>Chemical extraction</i>	10. Melting techniques
3. Education	21. <i>Nonferrous smelting</i>	11. Sand casting
4. Statistics and economics	22. Distillation	12. Permanent mold casting
5. Plant management	23. Electrolytic processes	13. Die casting
6. Industrial relations	25. <i>Nonferrous vacuum methods</i>	14. Centrifugal casting
7. Health and safety	26. Reduction by metals and gases	15. Precision investment casting
8. <i>Industrial pollution and waste disposal</i>	27. <i>Precipitation (replacement)</i>	16. Other methods
9. Research and development	28. Separation of metals	17. Pattern making
11. <i>Conservation of resources</i>	29. Amalgamation	18. Sand technology
12. Professional activities	1. <i>Decomposition</i>	19. Molding
13. Laws and legal actions	5. <i>Nonferrous melting and casting</i>	21. Coremaking
14. Literature techniques	6. <i>Isotope separation</i>	22. Rigging
15. Audio-visual techniques		23. Pouring
B — Ore and Raw Material Preparation	D — Iron and Steelmaking	24. <i>Casting cleaning</i>
12. Mining methods	1. Blast furnace practice	25. Metallurgical control
13. Crushing, grinding and sizing	2. Openhearth practice	26. <i>Casting reclamation</i>
14. Concentration (beneficiation)	3. Bessemer process	F — Primary Mechanical Working
15. Roasting and calcining	5. Electric arc processes	21. Preparation for working
16. <i>Sintering processes</i>	6. Electric induction practice	22. Forging (hammer and press)
17. <i>Fuel distillation and gasification</i>	7. <i>Multistage processes</i>	23. Rolling
18. Processing of fuels	8. <i>Special processes</i>	24. Extrusion of rod, shapes and tube
19. Processing of refractories	9. Ingot casting	25. Swaging
23. Processing of scrap	10. <i>Oxygen furnace processes</i>	26. Tubemaking (except extrusion)
24. Water treatment	11. <i>Physical chemistry of iron and steel-making</i>	27. Drawing of rod and shapes
25. Production of industrial gases		

- 28. Wire drawing
 - 29. Finishing operations
- G — Secondary Mechanical Working
- 1. Presswork in general
 - 2. *Shearing*
 - 3. Stamping
 - 4. Drawing
 - 5. Impact and cold extrusion
 - 6. Bending
 - 9. Stretch forming
 - 10. Heading
 - 11. Roll forming
 - 12. Thread rolling
 - 13. Spinning
 - 14. Bulging
 - 16. Die sinking
 - 17. Machining, *mechanical*
 - 18. Grinding
 - 19. *Finish machining*
 - 22. Flame and arc cutting
 - 23. *Finish forming*
 - 24. *Special machining methods*
- H — Powder Metallurgy
- 10. Powder production
 - 11. Powder properties
 - 12. Mixing and blending
 - 14. Molding and compacting
 - 15. Sintering
 - 16. Post-sintering operations
 - 17. *Fiber metallurgy*
- J — Heat Treatment
- 21. Homogenization
 - 22. Austenitizing
 - 23. Annealing
 - 24. Normalizing
 - 25. Patenting
 - 26. *Quenching*
 - 27. Precipitation hardening
 - 28. Case hardening
 - 29. Tempering
 - 1. *Residual stress control*
 - 2. Heating and cooling methods
 - 3. *Other carbon diffusion processes*
 - 4. *Surface alterations*
 - 5. *Hardenability*
- K — Assembling and Joining
- 1. Arc welding
 - 2. Gas welding
 - 3. Resistance welding
 - 4. Thermit welding
 - 5. Forge welding
 - 6. Other welding processes
 - 7. Soldering
 - 8. Brazing
 - 9. *Process control*
 - 11. Bonding metals to nonmetals
 - 12. Adhesive joining
 - 13. *Mechanical assembling*
- L — Cleaning, Coating and Finishing
- 10. Mechanical cleaning and polishing
 - 12. Chemical cleaning and polishing
- M — Metallography, Constitution and Primary Structure
- 20. *Specimen preparation*
 - 21. Microscopy
 - 22. Diffraction methods
 - 23. Other techniques
 - 24. Phase diagrams
 - 25. *Atomic and molecular* structures
 - 26. Crystal structures
 - 27. *Primary* microstructures
 - 28. Macrostructure
- N — Transformations and Resulting Structures
- 1. Diffusion
 - 2. Nucleation
13. Electrochemical cleaning and polishing
14. Chemical *conversion* coating
15. Diffusion coating
16. Dip coating
17. Electroplating
18. Electroforming
19. *Anodic coating*
21. Cathodic oxide coating
22. Cladding or bonding
23. Metal spraying
24. *Coating by weld-deposition*
25. *Coating by vapor-deposition*
26. Painting and organic coating
27. Ceramic coating (vitreous enameling)
28. *Coating by chemical reduction*
29. *Other methods*

3. Grain growth
 4. Recovery
 5. Recrystallization
 6. *Diffusionless transformations*
 7. Precipitation
 8. Austenite formation and decomposition (Fe-C system)
 9. Other eutectoid reactions
 10. Superlattice formation
 11. Other solid-state reactions
 12. Liquid-solid reactions
 14. Liquid-liquid reactions
 15. Gas-solid reactions
 16. Gas-liquid reactions
- P — Physical Properties
10. Mass and volume
 11. Thermal
 12. Thermodynamics and kinetics
 13. *Surface properties*
 15. Electrical and electrochemical
 16. Magnetic
 17. *Radiation properties (electromagnetic)*
 18. *Nuclear*
- Q — Mechanical Properties and Tests
21. Elastic properties.
 22. Anelastic properties
 23. Plastic properties
 24. Plastic deformation *mechanism*
 25. Stresses
 26. Fracture *mechanism*
 27. Tension test
28. Compression test
 29. Hardness test
 1. Torsion test
 2. Shear test
 3. Creep test
 5. Bend test
 6. Impact test
 7. Fatigue test
 8. Damping test
 9. Wear tests
 10. *Other tests*
- R — Corrosion
1. Mechanism of corrosion
 2. Corrosive effects
 3. Atmospheric corrosion
 4. Waters
 6. *Chemicals, inorganic*
 7. Chemicals, organic
 8. Soils
 10. Preventive measures
 11. Tests
- S — Inspection and Control
10. Sorting and identification
 11. Composition analysis
 12. Statistical *analysis and control*
 13. Flaw detection
 14. Size, *thickness and mass* measurement
 15. Surface roughness
 16. Temperature measurement and control
 18. Process control and *measurement*
 19. Radiation detection and *measurement*
21. Miscellaneous service testing (life testing)
 22. Standard practice and specifications
 23. *Calibration*
- T — Metal Products and Parts
21. Automotive
 22. Marine
 23. Railroad
 24. Aircraft
 26. Structures
 29. Chemical processing
 1. *Electrical and electronic components*
 2. *Ordnance material*
 3. Agriculture
 4. Civil engineering
 6. *Tools*
 7. Machine parts
 9. Arts and crafts
 10. Consumer goods
 11. *Nuclear reactor components*
- U — Allied Fields
1. Physics
 2. Chemistry
 3. Ceramics
 4. Mathematics (statistics and nomographs)
 5. Geology
 6. Mineralogy and petrography
 7. Engineering
 10. *Applied mechanics*
 11. Stress analysis and elasticity
 12. *Biology*

W — Plant Equipment

- | | | |
|---|--|--------------------------------------|
| 10. Plant installations | 27. Heat treating furnaces | 28. Elastic and anelastic properties |
| 11. Power generation and distribution | 28. Other heat treating equipment | 29. Mechanical properties |
| 12. Materials handling and transportation | 29. Welding, brazing and soldering | 1. Corrosion |
| 13. General service equipment | 1. Mechanical joining | 2. Radiation detection |
| 15. Coal and ore preparation | 2. Mechanical cleaning and finishing | 3. Micrography |
| 16. Fuel distillation and gasification | 3. Chemical and electrochemical cleaning and finishing | 4. Metallography |
| 17. Furnaces and equipment for primary metallurgy | 4. Other cleaning and finishing | 5. Photography and other visual aids |
| 18. Melting and refining furnaces and equipment | X — Instrumentation — Laboratory and Control Equipment | 6. Sorting and identification |
| 19. Casting | | 7. Atmosphere and humidity |
| 20. Ingot and slab preparation | 20. Size, thickness and mass | 8. Flaw detection |
| 22. Forging | 21. Composition | 9. Temperature |
| 23. Rolling | 23. Surface properties | 10. Electrical energy |
| 24. Forming and shaping | 24. Thermal properties | 11. Magnetic fields |
| 25. Machine tools | 25. Electrical properties | 12. Pressure |
| | 26. Magnetic properties | 13. Rate and time |
| | | 14. Computers |
| | | 15. Communication |

Materials Index

- | Common Elements and Their Alloys | |
|----------------------------------|----------------------|
| Ag—Silver | Cu—Copper |
| Al—Aluminum | Fe—Iron |
| Be—Beryllium | Mg—Magnesium |
| C—Carbon | Mn—Manganese |
| Cb— <i>Columbium</i> | Mo—Molybdenum |
| Co—Cobalt | Ni—Nickel |
| Cr—Chromium | Pb—Lead |
| | Si—Silicon |
| | Sn—Tin |
| | Ta— <i>Tantalum</i> |
| | Ti— <i>Titanium</i> |
| | U— <i>Uranium</i> |
| | V— <i>Vanadium</i> |
| | W— <i>Tungsten</i> |
| | Zn—Zinc |
| | Zr— <i>Zirconium</i> |

Materials Groups		SGB — Materials by Properties and Applications (Group B)		j. Abrasives k. Other	
EG—Element Groups (individual elements coded by chemical symbol)		n. <i>High nuclear cross-section</i>		RM — <i>Industrial Raw Materials</i>	
ST—Steel		p. <i>Low nuclear cross-section</i>		g. <i>Industrial gases (non-elemental)</i>	
CN—Carbon Steel		q. <i>Hard materials</i>		h. <i>Refractories</i>	
AY—Low-Alloy Steel (5% Alloy Max.)		r. <i>Low-temperature</i>		j. <i>Solid fuels</i>	
SS—High-Alloy Steel (Including Stainless and Heat Resisting)		s. <i>Structural</i>		k. <i>Liquid fuels</i>	
TS—Tool and Die Steel		a. <i>High-strength</i>		m. <i>Gaseous fuels</i>	
CI—Cast Iron			NM — <i>Nonmetallic Materials</i>	n. <i>Ores</i>	
SGA—Materials by Properties and Applications (Group A)		a. <i>Inorganic compounds</i>		p. <i>Scrap</i>	
SGB—Materials by Properties and Applications (Group B)		b. <i>Organic compounds</i>		q. <i>Fluxes and slags</i>	
NM—Nonmetallic Materials		c. <i>Catalysts</i>		AD — <i>Addition Agents</i>	
RM—Industrial Raw Materials		d. <i>Polymers</i>		n. <i>Alloying agents</i>	
AD—Addition Agents		e. <i>Minerals (nonmetallic)</i>		p. <i>Structure-controlling agents</i>	
		f. <i>Ceramic materials</i>		q. <i>Agents imparting specific properties</i>	
		g. <i>Coatings (nonmetallic)</i>		r. <i>Deoxidizers</i>	
		h. <i>Lubricants and oils</i>		s. <i>Oxidizers</i>	
				a. <i>Desulphurizers</i>	
		Common Variables Index			
1. <i>Manufacturing and Test Methods</i>		17. <i>Cold</i>		10. <i>Effect of composition</i>	
2. <i>Use of equipment</i>		18. <i>Electrical</i>		11. <i>Effect of temperature</i>	
4. <i>Test methods</i>		19. <i>Induction</i>		12. <i>High temperature</i>	
10. <i>Specimen preparation and design</i>		20. <i>Hydraulic</i>		13. <i>Low temperature</i>	
11. <i>Continuous</i>		21. <i>Pneumatic</i>		14. <i>Effect of heat treatment</i>	
12. <i>Batch</i>		22. <i>Mechanical</i>		15. <i>Effect of aging</i>	
13. <i>Cyclic</i>		23. <i>Vacuum</i>		16. <i>Effect of chemical environment</i>	
14. <i>Acid</i>		24. <i>Ultrasonic</i>		17. <i>Effect of irradiation</i>	
15. <i>Basic</i>		2. <i>Influencing factors</i>		3. <i>Influencing factors</i>	
16. <i>Hot</i>		9. <i>Effect of grain size</i>		16. <i>Effect of stress</i>	

17. Effect of time or rate
18. Effect of deformation or strain
19. Effect of impurities
20. Effect of prior history
21. Effect of prior structure
22. Effect of orientation
23. Size and shape effects
24. Effect of pressure
4. Metal Forms, Wrought
 1. Forgings
 2. Blooms and billets
 3. Sheet, strip and plate
 5. Bars and rods
 6. Foil
 7. Shapes (I-beams, rails, etc.)
 8. Extrusions
 9. Stamping
 10. Tubing and pipe
 11. Wire
 12. *Wool*
5. Metal Forms, Cast
 9. Ingots
 10. Sand Castings
 11. Die castings
 12. Investment castings
 13. Permanent mold castings
 14. Semi-permanent mold castings
 15. Centrifugal castings
 16. Chill castings
6. Metal Forms, Powder
 17. Pellets and shot
 18. Loose powders
19. Cemented carbides
20. Metal ceramics
21. Porous metal *parts*
22. Structural parts
23. *Briquettes*
7. Metal Forms, Assembled
 1. Welded
 2. Brazed and soldered
 3. Riveted
 4. Bolted
 5. Screwed
 6. Pressed fits
 8. *Bonded*
 9. *Composite*
8. Metal Forms, Coated
 12. *Electroplated*
 15. Dip coated
 16. Clad or bonded (*metal to metal*)
 17. Sprayed
 18. Hard surfaced
 19. Chemically coated
 20. Painted
 21. Enameled
 22. Vapor coated
9. Defects
 17. Pipe
 18. Porosity or blowholes
 19. Segregation and inclusions
 20. Flakes
 21. Surface defects (seams, laps, etc.)
 22. Cracks (quenching, grinding, corrosion)
10. Type of literature
 1. Theory
 2. Research
 3. Original techniques
 4. Review
 5. Plant description
11. Form of literature
 9. Patent
 10. Trade literature
 12. Specification
 13. Report
 14. Book
 15. Bibliography
 16. *Translation*
 17. *Glossaries, definitions, directories*
 18. *Data books and handbooks*
 19. *Trade names*
 20. *Microfilm and microcards*
12. Language
 17. French
 18. German
 19. Russian
 20. Italian
 21. Oriental
 22. Swedish
 23. Spanish
 24. *English*
13. Geography
23. Overheating
24. Distortion
 1. Impurities
 2. *Scale*

- | | | |
|---|---|---------------------------------------|
| 14. <i>States of Metals</i> | 21. <i>A.P.I.</i> | 12. <i>Atomic</i> |
| 9. <i>Minerals</i> | 22. <i>A.M.S.</i> | 13. <i>Solar</i> |
| 10. <i>Liquid metals</i> | 23. <i>A.S.M.E.</i> | 17. <i>General Concepts</i> |
| 11. <i>Single crystals</i> | 24. <i>Government</i> | 1. <i>Design</i> |
| 12. <i>Thin films</i> | 4. <i>I.S.O. (International Organization for Standardization)</i> | 2. <i>Performance characteristics</i> |
| 13. <i>Isotopes</i> | 5. <i>B.S. (British)</i> | 3. <i>Cost</i> |
| 15. <i>Colloids</i> | 6. <i>D.I.N. (German)</i> | 4. <i>Quality</i> |
| 16. <i>Amorphous</i> | 7. <i>U.N.I. (Italian)</i> | 5. <i>Tolerances</i> |
| 17. <i>Solid solutions</i> | 8. <i>C.S.A. (Canadian)</i> | 6. <i>Models</i> |
| 18. <i>Compounds (oxides, hydrides, carbides, etc.)</i> | 9. <i>N.F. (French)</i> | 7. <i>Applications of Materials</i> |
| 19. <i>Heat treated</i> | 10. <i>G.O.S.T. (Russian)</i> | 18. <i>Plant Practice</i> |
| 15. <i>Material Designations (Specifications and Standards)</i> | 11. <i>Other</i> | 17. <i>Layout</i> |
| 17. <i>A.S.A.</i> | 16. <i>Energy sources</i> | 19. <i>Construction</i> |
| 18. <i>A.S.T.M.</i> | 9. <i>Chemical</i> | 20. <i>Erection and installation</i> |
| 19. <i>S.A.E.</i> | 10. <i>Mechanical</i> | 21. <i>Maintenance</i> |
| 20. <i>A.I.S.I.</i> | 11. <i>Electricity</i> | 22. <i>Repair</i> |
| | | 23. <i>Lubrication</i> |
| | | 24. <i>Automation</i> |

Chapter 6

THE PEEK-A-BOO SYSTEM—OPTICAL COINCIDENCE SUBJECT CARDS IN INFORMATION SEARCHING*

W. A. WILDHACK and JOSHUA STERN

National Bureau of Standards, Washington, D. C.

Introduction

The "Peek-a-boo" principle, which is treated in detail in this chapter, is illustrated in Figure 6-1. The perforated cards shown represent subject headings or index terms. If a document being indexed relates to a particular index term, the card representing that term contains a hole at a position dedicated to that document. Thus each card representing terms relating to the document will have a hole at the identical position for that document; all other cards will have no hole at that position. The collection of holes on a given card identify directly all documents relating to the corresponding subject. If two or more cards are superimposed, holes which are not obscured, i.e., those which appear on all of the superimposed cards, identify documents each of which relate to all of the subjects represented. The document serial number is represented by the location of the hole, usually read by means of an imprinted pattern on the card or with a transparent overlay. A detailed description of the mechanical operation of the system will be found on pages 136-141.

The "Peek-a-boo" search mechanism was designed for a specific application and the features of the mechanism were dictated by the needs of that application¹. Many other ways of applying the principle have been devised which are advantageous under given circumstances. The principle itself appears to be quite old. The earliest application known by the authors was for identification of birds and was the basis of a patent issued in 1915.² Shortly thereafter, application to a number-guessing game was described in a publication of very limited distribution^{3, 4}. Identification of minerals

* Copyright excluded.

¹ Wildhack, W. A., Stern, J., and Smith, J., "Documentation in Instrumentation," *Am. Doc.*, 5, 223-37, Oct. 1954.

² Taylor, H., Selective device, U. S. Patent 1,165,465, December 28, 1915.

³ Gerardin, A., *Sphinx-Oedipe*, 11, 68-70 (1916).

⁴ Kraitchich, Maurice, *Mathematique des jeux ou recreations mathematiques*,

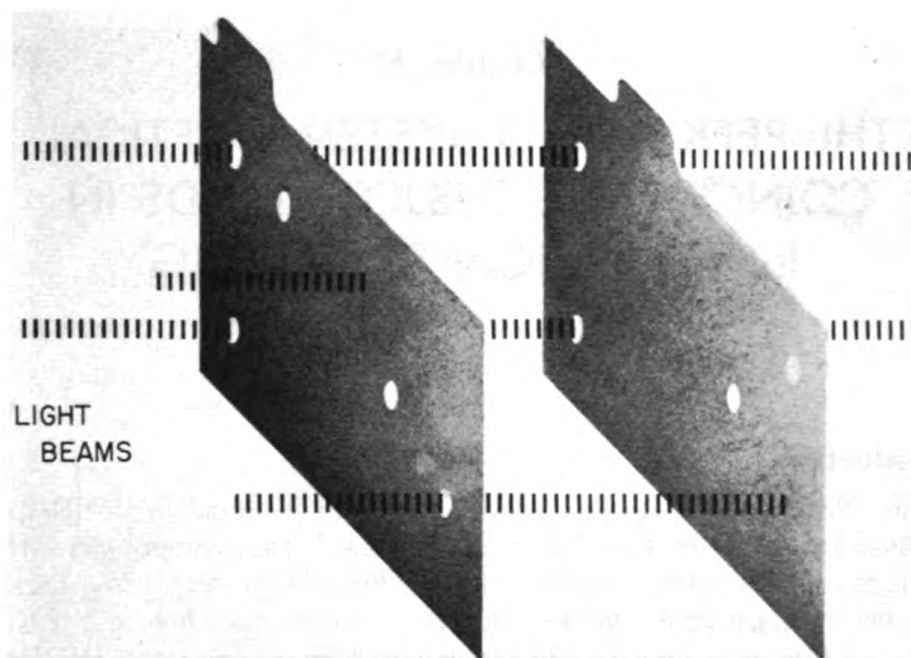


Figure 6-1. "Peek-a-boo" system, principle of operation. (Light beams are passing from left to right.)

by application of the principle is described by Gray⁵ and by Donnay^{6, 7}. An elaboration of the principle, disclosed in a patent issued in 1920⁸, deals with means for avoiding the need to remove cards from the file for reading. As described in the patent, this is accomplished by interspersing fully perforated reading columns between the normal data-punched columns. To read the file, the subject cards to be compared are displaced with respect to the bulk of the file to cause the data columns to coincide with the intermediate, fully-perforated columns of the rest of the cards. All the data positions on the cards being compared are thus exposed to view through the perforations in the reading columns, unobstructed by the cards not involved in the search. This patent cited no prior art and the patent office actions did not disclose prior art. Similar systems with addition of photo-electric read out, for use in telephony, are described by Myers⁹ and by

Ed. 1, Brussels, 1930; Ed. 2, Brussels, 1953. Revised editions in English, Ed. 1, New York, 1942; Ed. 2, New York, 1952, London 1943.

⁵ Gray, C. J., A new method of using the physical characteristics of minerals for their identification, *Trans. Geol. Soc. S. Afric*, **23**, 114-117 (1920).

⁶ Donnay, J. D. H., *Ann. soc. geol. Belg.*, **59**, B 250 (1936).

⁷ Donnay, H. D. H., *Am. Mineralogist*, **23**, 91 (1938).

⁸ Soper, H., Means for compiling tabular and statistical data, U. S. Patent 1,351,692, Aug. 31, 1920.

⁹ Myers, O., Electromechanical translator, U. S. Patent 2,558,577, June 26, 1951.

Gent¹⁰. The prior art described above was not cited by either the inventor or examiner in these patents. Another photoelectric read out in which the various hole positions are scanned one at a time, so that only one photocell is required, is described by Drillick¹¹. The principle has been applied to searching personnel records as described in a French patent issued in 1923¹². Application of the principle to patent files has been made by Batten¹³.¹⁴, ¹⁵. This application used hand-punched and hand-manipulated cards with a relatively low density of punching. An application of the principle which has been commercially available abroad was devised by Cordonnier¹⁶.¹⁷, ¹⁸. This system, believed to be based on the French patent cited above¹², is trade-named "Selecto". It uses in one form a card 15 x 21 cm with a capacity of 12,500 documents. The card is pre-printed with a grid identifying the hole positions so that no read-out device is needed. Holes are approximately 0.7-mm diameter. Cards of larger and smaller capacities are available. Also commercially available is a French system called "Sphinxo". The Sphinxo card measures approximately 5 $\frac{5}{8}$ x 10 $\frac{1}{2}$ inches and has a document capacity of 1000¹⁹. The authors have no information on the connection, if any, between this system and the disclosure cited above³, published in the Journal *Sphinx-Oedipe*. A card 8 $\frac{1}{4}$ x 11 $\frac{5}{8}$ inches, with 7000 document spaces, is manufactured in Germany under the name "Ekaha"²⁰. Also of German manufacture are "Sichtlochkarten"²¹. These

¹⁰ Gent, E. W., et al, Electromechanical translator, U. S. Patent 2,668,877, February 9, 1954.

¹¹ Drillick, J. H., Fast access to punched card data file, *Product Eng.*, **22**, 176-178, Oct. 1951.

¹² Lieber, Henri, French Patent 565,745, Nov. 27, 1923.

¹³ Batten, W. E., A punched card system of indexing to meet special requirements. Report of the 22nd Conference, ASLIB (4 Queens Gate, London, W.8) pp. 37-39 (1947).

¹⁴ Batten, W. E., Specialized files for Patent Searching, Casey, R. S. and Perry, J. W., "Punched Cards, Their Application to Science and Industry," Chapter 10, pp. 169-181, New York, Reinhold Publishing Corp., 1951.

¹⁵ DeGorter, B., The indexing of British patents on plastics. I.C.I. Ltd., Plastics Div., Intelligence Section, (C4/BdeG/JH) 18.2.47 (Hollerith).

¹⁶ Cordonnier, G., Classification, classement, rangement et selection Revue Mensuelle de l'organisation, Comite National de L'Organisation Francais, 57 Rue de Babylone, Paris 7e, Avril-Juillet 1951.

¹⁷ DeMamantoff, N., Chef du Service, Le Centre de Documentation du Centre Nationale de La Recherche Scientifique, Utilisation du Systeme Cordonnier. (1953) (16 rue Pierre-Curie, Paris 5e).

¹⁸ Schurmeyer, Dr. Walther, Selecto—Ein neues Auswahlssystem fur Dokumentation, Nachrichten fur Dokumentation, **3**, #1, 33, Mar. 1952.

¹⁹ Detectri, 68 rue de Richelieu, Paris 2e, France.

²⁰ Edler & Krische, Hannover, Germany. Ekaha cards (7000 docs. 8 $\frac{1}{4}$ x 11 $\frac{5}{8}$ inch).

²¹ Allform Buro-organization GmbH, Brandenburgische Strasse 27, Berlin W 15.

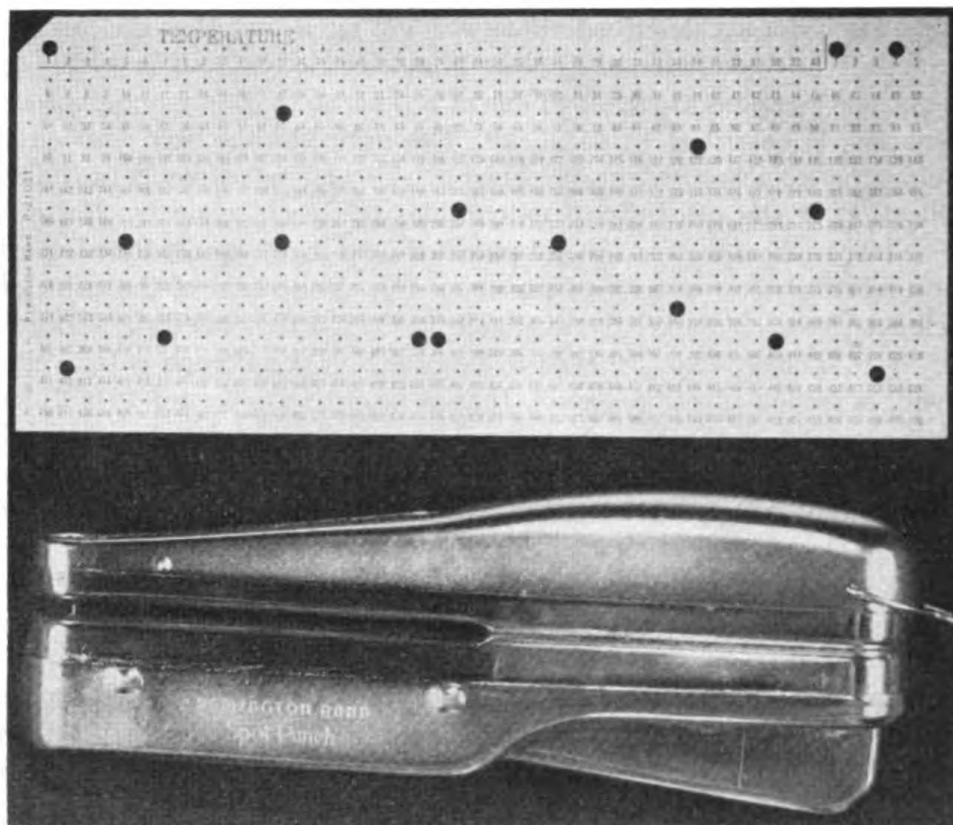


Figure 6-2. "Vicref" card for hand punching. Capacity 500 documents.

are available in approximately 5 x 8 inch size with capacities of 2000 and 6000. Application of this card and of the Ekaha cards to pharmaceutical indexes has been described in Reference 22.

Accounting-machine cards, and the punches available for such cards, have been used in applying the Peek-a-boo principle. Robinson²³ has used such cards to prepare stencils for solving mathematical equations. An application which uses a specially printed card has been devised by Seely²⁴, and named "Vicref"—for Visual Gross Reference. The card,* illustrated in Figure 6-2, is pre-punched with pilot holes and pre-printed with document numbers. When used with a suitable hand punch, the pilot holes ensure perfect registry of the larger document-identification holes. The card has a capacity of 500 documents. Additional sets are envisaged to increase the file to multiples of this number and provision is made for identification of up to 40 sets by a series of hole positions reserved for this purpose. The

²² Adler, Fritz H., *Pharm. Ind.*, 19, 170-172 (1957). Allform and Ekaha cards in pharmaceutical indexes.

²³ Robinson, R. M., *Stencils for solving $x^2 = 2 \pmod{m}$* , U. of Cal. press 1940.

* Obtainable from Sperry-Rand Corporation.

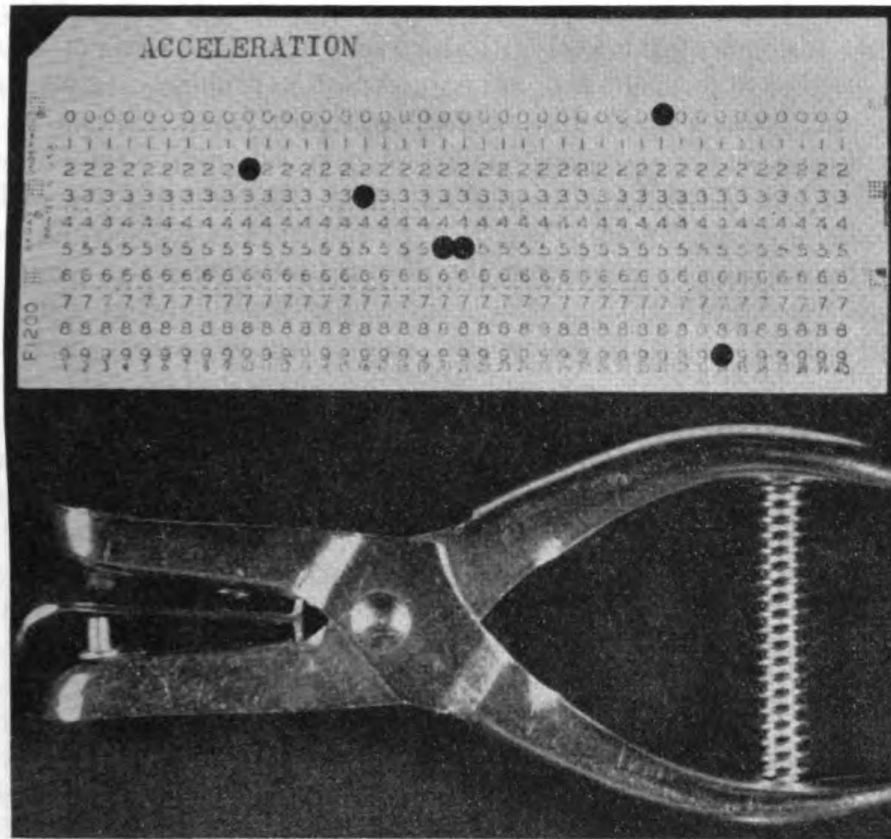


Figure 6-3. Samas card for hand punching, capacity 400 documents.

authors have experimented also with the smaller accounting-machine cards* shown in Figure 6-3a, using an ordinary "conductor's" punch to make the perforations. Recently introduced commercially in this country are "Termatrex" cards and equipment in two capacities, 15,000 and 40,000 documents²⁵. More elaborate implementations of the Peek-a-boo principle have also been made. One of these involves keyboard selection of the cards to be punched or read, simultaneous drilling of all selected cards, and random return to the file²⁶.

The information indexing and searching system to be described here was developed at the National Bureau of Standards as part of a program of research in methods of measurement and in improvement of scientific instruments.

* Obtainable from Underwood Corporation.

²⁴ Seely, J. S., Vicref (visual cross reference), private communication 1952.

²⁵ Documentation Incorporated, 2521 Connecticut Ave., N. W., Washington 8, D. C.

²⁶ Miller, Eugene, Final Report to the National Science Foundation on the Matrex Indexing Machine, Jan. 1957, 13pp.

The program envisaged three kinds of activity: (1) research and development in instrumentation; (2) critical surveys of various areas of instrumentation; and (3) consultation on instrumentation problems. In planning the program it was evident that all three activities would rely heavily on documentation, defined as the handling of recorded information. It was also evident that instrumentation literature is poorly organized and is hard to search, being found as a part of all branches of science and technology. Studies toward improving the documentation of instrumentation literature were therefore undertaken involving the following elements: locating sources, selecting references, identifying references, indexing, coding, filing, storing, searching, and retrieving. These studies of the general problems, initiated in 1950, revealed the difficulties of conventional hierarchical classifications as the basis for indexing, and led to recognition of the essential "multi-dimentional" or "multiple-independent-aspect" nature of information, and thus to consideration of the "multi-dimensional" or "multi-aspect" approach to the indexing problem.

The information indexing system which finally evolved is based on multi-aspect classification of document content and Peek-a-boo searching.

Fundamental Considerations

The key element in establishing an information file is the indexing process. This should result in the assignment of class designations which function to disclose to the searcher those documents in which he may expect to find the information he is seeking. Logical steps in the indexing operation are: (1) establishment of index classes to which a document may be said to belong on the basis of its content; (2) assignment of names, or class designations to the document classes thus established; (3) recognition of that content information which is to be represented in the index; (4) assignment of the document to all appropriate index classes on the basis of content so recognized. A brief review of the general problems of indexing will provide background for a description of the system adopted for indexing, specifically as applied to the literature of instrumentation.

The following may be set forth as the specifications to be met by an indexing procedure:

- (1) The index must be capable of representing all information which is to be retrievable in the future.
- (2) The "language" of the index needs to be common to both the indexer and the searcher.
- (3) The classes of documents established by the index must be small enough to be acceptable to the searcher.
- (4) All classes of documents required by the searcher must be present in the index either specifically or as obvious members of more generic classes which are themselves small enough to be acceptable as the search product.

Each of these specifications implies anticipation of the desire of the searcher. To meet the first specification, the capability of representing all information which is to be retrievable, the index must have terms which subsume, or are related (logically or by arbitrary assignments) to all the types of information to be retrieved. The ability of the indexer to recognize all such information is assumed, since no index can compensate for failure on this point.

A large fraction of the problems of indexing relates to the second specification, that of providing an index language common to indexer and searcher. The ambiguity of language in general is a basic difficulty for which a remedy is yet to be found. A first step is the rigorous definition of all terms used in the index. Such definitions may well be arbitrary, provided they are unambiguously understood by both indexer and searcher. More specific are the problem of synonyms, and the problem of generic relationships. In addition to the ordinary synonym, for example, "aerial" and "antenna," synonymous expressions and word-order synonyms must be recognized. Examples of the synonymous expression are "force" and "rate of change of momentum," and "pressure" and "force per unit area." Examples of the word-order synonym are "electric motors" and "motors, electric," or "radar for air navigation" and "air navigation by radar." The problems presented by the word-order synonym and by the synonymous expression are rapidly compounded as the number of words in the index subject heading increases.

The problem of generic relationships among index terms is related to the problem of synonyms. The expression which is synonymous with a single word may be regarded as a special case of generic relationships. Since the whole is the sum of its parts, the existence of generic relationships offers alternative, and hence ambiguous, use of the index terms in searching (i.e., a given piece of information may be sought *via* a general index term which describes it or *via* the combination of a number of terms which together describe the information). This aspect of the generic relationship is classed here as a language problem since its existence results in an uncertainty in the mind of the searcher in regard to the choice of terms made by the indexer. Aspects of the generic relationship problem which influence the ability of the index to meet specifications (3) and (4) are treated below.

These language problems have traditionally been treated in the following way:

(a) The index language is formalized so that the index vocabulary is strictly limited.

(b) Synonyms are eliminated or cross referenced, synonymous expressions being treated in the same way as simple synonyms.

(c) Generic relationships are prevented from causing trouble by requiring that only established terms or established combinations of terms be

used. In this way it is possible to provide a fixed hierarchy so that all recognized generic relationships are explicit, permitting the searcher to select the most specific index term which meets his needs.

(d) The word-order synonym is treated by cross reference in the same manner as are other synonyms. No more than a minute fraction of the problem can be dealt with in this way because of the tremendous number of such synonyms which exist.

These remedies for the language problem are limited severely by the ambiguity of language which makes impossible the establishment of a universally understood index vocabulary which will not be eroded by time. Furthermore, while the techniques mentioned contribute in good measure to the solution of the language difficulty, they are not entirely compatible with the other requirements of indexing, particularly specifications (3) and (4).

Specifications (3) and (4) deal with establishment of the classes of documents which are to be the product of the information search. These requirements of an index are basic. The total collection (representing the most general class of documents) must be broken down into classes small enough to be searched further, without additional help from the index. (The existence of auxiliary specific indexes influences the acceptable size of the product class. For example, if the document is a book, its own index will carry the burden of searching within the document.) The size of the search product which is acceptable varies widely with the circumstances of the search so that it is not possible to correlate the *kind* of information to any "acceptable" size. Usually arbitrary bounds are set on the kind of service to be performed by a given index. When one is seeking a specific datum for immediate application in conjunction with other data only the most specific class will be acceptable. For example, in seeking information on the melting point of a particular organic compound, a class which contained all documents in which properties of organic compounds were discussed would be very unsatisfactory. For this kind of search a handbook type of index system which relates the specific property and the specific compound is necessary. On the other hand, a search designed to disclose the state of the art of measurement of flow of liquids would tolerate—or, in fact, demand—a very large search product. The kind of service to be fulfilled by the index needs to be anticipated and carefully defined. If one considers the establishment of index classes without anticipating future meaningful reference questions, the number of valid classes which can be defined approaches infinity. Even when limits are placed on the system by anticipating the kinds of questions which will be asked, it is still necessary to provide a larger number of search-product classes than can be recognized explicitly. Attempts to recognize a sufficient number of meaningful specific document

classes magnify enormously all aspects of the language difficulties. The cross-reference technique quickly becomes inadequate. An approach to indexing which can provide the large number of classes needed is that which has been referred to above as the multi-dimensional or multi-aspect approach, in which each search-product class is specified in terms of common membership of each document in a number of more general classes designated explicitly by general index terms. The number of specific document classes which can be formed in this way far exceeds the number of general classes (index terms) which are explicitly established. For example, with a vocabulary of 1000 general terms the number of specific classes which can be generated by combining up to 10 terms in every possible way is approximately 10^{22} . Of course many of the classes established in this way will not relate to any actual document and some may include contradictory elements and hence will not have even potential utility.

The principle of combination thus gives the very large number of classes needed without requiring that each be separately recognized in advance. Whether the classes which can be defined in this fashion include the ones which are needed for search purposes remains to be seen. If the class names recognized explicitly in the index consist exclusively of single words, no word-order problem can arise. It is usually not desirable to do this fully.

This approach implies abandonment of some of the techniques listed above which have been used in the past to control the "language" problem as outlined above. The principle of *formation of specific classes at the time of search by combination of a number of general class designations* is not compatible with rigid formalization of the search language. This in turn means that the synonym problem is increased in severity by the variety of ways in which a given document can be specified by combinations of index elements. These factors, particularly when combined with the ambiguities inescapable in any system, and with the uncertainty whether the increased number of classes implies a significant increase in the number of *useful* classes, raise questions as to the net gain derivable from the newer techniques. Nonetheless, it is possible to conclude that there is a net gain, on the basis of specific advantages, even though these too may not be susceptible to quantitative evaluation. If we regard an index with no provision for generating new classes by combination of recognized classes as one extreme of a spectrum of index systems, and the system in which recognized classes are never used alone, but are always combined to form the end-product class sought, as the other extreme, then it is evident that all systems fall somewhere within this spectrum. The problem of designing an index which uses combination techniques is that of finding the optimum position along this spectrum. This assumes that effective techniques can be found to solve the mechanical problem of performing the operation of combining classes to

form other classes and to isolate the resulting classes. The following discussion deals with these techniques and in particular with the techniques to which the name "Peek-a-Boo" has been given.

Search Mechanisms

Having devised a search classification or index which depends upon the common association of a document with a number of different index terms to identify that document, it remains to select mechanisms for recording and detecting such associations. A wide variety of such mechanisms have been devised, ranging from simple, hand-manipulated cards to the most complex electronic computers. Many more such devices have been proposed or can be conceived.

By and large, devices for this purpose fall into two classes: those in which information identifying a document is associated, inseparable physically, with all the index information relating to the document, and those in which symbols for all documents relating to a given item of index information are associated, in a physically inseparable fashion, with each such index term. If the physical entity accomplishing the necessary association is a card, we have on the one hand "document cards" each of which carries all the pertinent subject headings and, on the other hand, "subject cards" each carrying all of the pertinent document identifications. The same situation holds if we substitute for the card, as the medium for associating index and document, any of the other means which have been used or proposed such as punched tape, photographic film, magnetic wire records, magnetic tape records, static magnetic memories, superconductive memories, etc.

It is convenient to classify indexing systems in this way—as document-card type or subject-card type—because with each of the two classes are associated important performance characteristics which need to be weighed in terms of a specific application to estimate its advantages and disadvantages.

A system characteristic of prime importance is the nature of the *presorts* permitted—a sort being defined for this purpose as the physical formation or isolation of a class as opposed to the mere definition of that class, and a presort being such a sort made at the time of filing rather than at the time of search. No matter how fast our mechanisms can search a collection of cards, film frames, or magnetic states, an additional increment of speed is obtainable if only a portion of the collection needs to be searched. This principle is, of course, the controlling one in the classic hierarchical approach to indexing which postulates "a place for everything, and everything in its place." All presorts which are based on setting up preferred groupings necessarily require anticipation of the search questions which will be asked of the index in order to select from the vast number of possible

groupings those which are to be preferentially presorted. It is therefore a matter of intuition, or guesswork, to predict the nature, extent, and importance of future failures. In systems which use the search principle of combining general indexes at the time of search to generate specific indexes, it is possible to make one kind of presort, as defined above, without requiring arbitrary preference. This is not because the presort is a perfect one but because it adds no additional imperfection beyond that which has already been accepted in the system. Given a set of general indexes which are to be combined for more specific searching, it is always possible to presort on the basis of each of the general indexes without introducing any new ambiguity into the future search. The reason for this lies quite simply in the fact that, in any case, every search must start by specifying at least one index. The presort on the basis of each index is the identical operation which must inevitably constitute the first step in the search. If such a presort has been made, then instead of searching the entire collection, it is only necessary to search further those items which have already been grouped in a compact form. The resulting saving of search time can be decisive. The nature of this presort will become clearer when a particular system is described.

This susceptibility to an unambiguous presort of this kind is characteristic of the class of search systems in which a medium, such as a card, is used to associate each index term with all the documents to which it pertains. For convenience we will henceforth refer to this class as the subject card class and to its converse as the document-card class, keeping in mind that the considerations involved are not limited to card systems.

A second characteristic, relatively less important, is the ability of a system to yield subject information as a direct product of a search—rather than a set of serial-number citations. This capability is a direct characteristic of the document-card class. It is possible to make a subject-card system perform in this manner as will be shown later.

Closely connected to the preceding is another characteristic relating to the adaptability of a system to replication and to current distribution of the information “cards.” Since indexing is perforce a document-by-document operation, the document-card system impresses no constraint on current distribution (analogous to the unambiguous subject presort of subject cards). The subject-card file, on the other hand, must await filling of the card before distribution of duplicates can be effected economically. It is feasible to distribute annual indexes in this way, replacing an incomplete set with an up-to-date one periodically. Alternatively, an initially complete index could be distributed, followed by currently distributed document analyses which the recipient would “punch in” to keep the index current.

Of great importance is the “open endedness” of the file for additions to,

and modifications of, the index vocabulary and for growth of the collection of documents. Document-card systems are limited in open endedness with respect to subject. A document-card system therefore demands very careful design of the subject index. Modifications tend to disrupt the system. Additions are usually severely limited by the capacity of the medium (card). The subject-card system has no such limitations, the addition of another subject index term requiring only an additional card. On the other hand, the subject card must necessarily have a limited capacity for document codes. Limitations on open endedness of either system are not absolute. In either system additional index terms in one case, or additional documents in the other, may be accommodated by using multiple cards. Additional cards used in this way are called "trailer cards." The use of trailer document cards involves the difficulty that selection may require interaction of subjects which appear on separate cards. This is not easily accomplished. Trailer subject cards do not encounter this difficulty since it is never necessary to compare the information on a card with information on any of its trailers.

In selecting a subject-card system rather than a document-card system as the search mechanism for use in the multi-aspect instrument reference system at NBS, the overriding considerations were: (1) the desirability of subject open-endedness for an experimental system, (2) the mechanical advantage in rapidity of search provided by the subject presort characteristic of the subject-card system. After some initial experimentation with subject cards on which were listed the serial numbers of pertinent documents, the subject-card system based on positional coincidence of perforations was selected as the most effective, mechanically simple, subject-card system known to us.

The Peek-a-Boo System. The name Peek-a-Boo was given to the particular subject-card system used for the instrumentation reference file as being noncommittal as to origin and descriptive of the operating principle which appears to have a rather long history (page 125), dating back at least as far as 1915².

The two cards shown in 6-1 represent index cards, with each one representing an index term. The position of each and every hole in a card is interpretable as a serial number of a corresponding document. If several such cards are superimposed, only those hole positions which are perforated in all of the cards will permit light to pass through. Hence, such illuminated positions identify documents with which all of the index terms represented by the superimposed cards are associated.

In adapting this principle to the instrument reference file, a card is assigned to each descriptive term in the index. The descriptive term is noted at the top of the card and the cards are filed alphabetically within each

category or sub-category. As an example, a document on "electromagnetic flowmeters for blood" would be recorded in the file on one card labeled "electromagnetic," on one labeled "flow," and on another labeled "blood" (or "liquid," if liquids were not further subdivided). The single card labeled "flow" would refer not to the one report of the example (Serial No. 10,924) but also to all other reports in the collection which have the word "flow" as one of their descriptive terms. The descriptive term card directs the searcher to the reference by means of the visible hole, one for each report to which that descriptive index term has been assigned. The location of this hole (row and column) identifies the serial number, and possibly the location, of the report. Thus the card labeled "flow" contains a hole in position 10,924; the card labeled "electromagnetic" also contains a hole in position 10,924, and the card "blood," or "liquid" also carries a hole in that position.

The Peek-a-boo instrumentation file is based on a card measuring approximately 5 x 8 inches with a capacity of approximately 18,000 ordered spaces for holes.

The Card. The material used for the card is a matter of some importance. Durability, dimensional stability, opacity, suitability for typewriting, punchability, permanence, and availability are factors which enter into the choice of material. Paper, which was used in the initial experimentation, was found deficient in durability and dimensional stability. This, however, is of importance only when a high density of information storage is desired. It is possible by appropriate design to overcome the difficulties resulting from dimensional instability¹ if one is willing to accept some operating limitations.

The Peek-a-boo cards more recently used at NBS are of vinylite plastic.* They are negligibly affected by humidity. Those qualities which require a longer time for their complete evaluation appear to be satisfactory so far. Typing requires a special ribbon which is, however, inexpensive and readily available. One corner is cut to ensure unambiguous orientation. Except for the index terms typed in the upper left corner, there are no markings on the card. Cards imprinted with a grid which would permit document numbers to be read directly were not obtainable at an acceptable price in the small quantities required. The need for precise printing registry and special card stock contribute to the high cost of preprinted cards. Figure 6-4 illustrates the cards and shows how the juxtaposition of two cards narrows a search based on two general terms.

Punching. The punch used to perforate a hole in the proper position corresponding to the serial number identifying a document is shown in Figure 6-5. It is designed to have sufficient precision to take advantage of

* Obtainable from Wassell Organization, Inc., Westport, Conn.

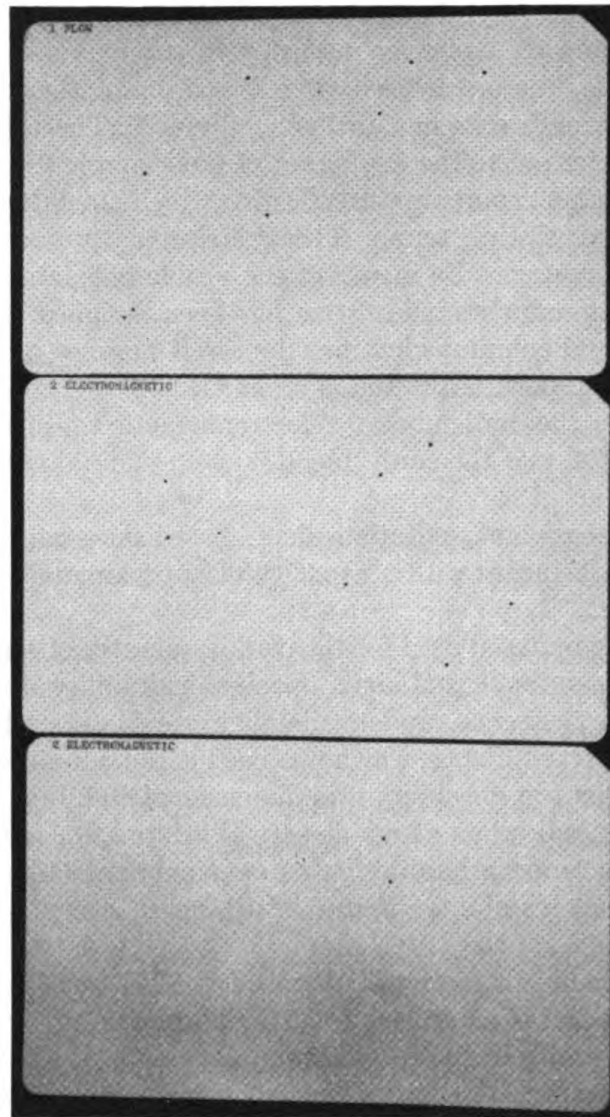


Figure 6-4. a) Peek-a-boo card showing documents relating to measurement of flow. b) Peek-a-boo card showing documents relating to electromagnetic principle of measurement. c) Superposition of the general "flow" card and the general "electromagnetic" card to specify documents relating to electromagnetic measurement of flow.

the dimensional stability of the card. The punch perforates holes approximately 0.6 mm in diameter in a square pattern spaced 1 mm center-to-center. In this way, an area 100 mm x 180 mm ($4 \times 7\frac{1}{4}$ inches approx.) on the 5 x 8 inch card accommodates 18,000 documents. Positioning is accomplished by two screws, one for each coordinate of motion. A single turn of a screw moves a fence one unit horizontally or vertically. The card to be punched is inserted so that two edges are located by the fence. Thus the

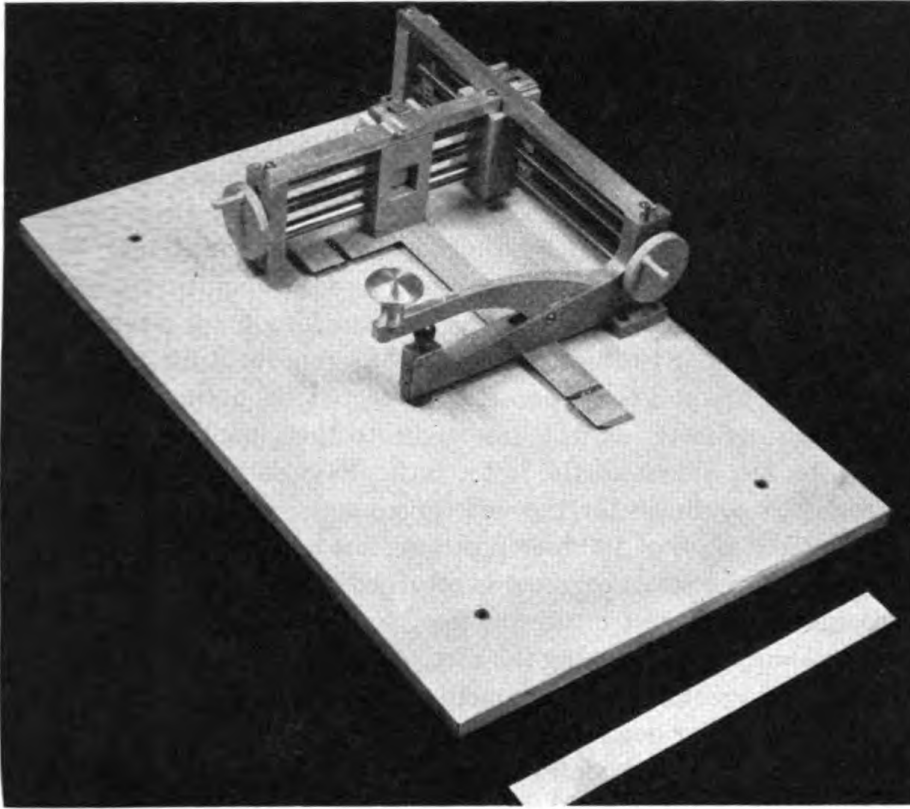


Figure 6-5. Index punch for Peek-a-boo cards.

position for successive documents is varied with respect to a fixed punch and die. The normal procedure in punching information is to proceed in the order of document serial number. In this procedure, adjustment of the punch to the proper position for successive documents is made by one turn of a knob. A detent eliminates any concern with visual setting of the knob by providing a "feel" for the correct position. If it is necessary to backtrack or to punch documents out of order, a split-nut release mechanism permits rapid traverse to the desired positions. Scales on the two axes identify the setting of the punch and provide a useful check from time to time.

Punching errors, when they are discovered, can be corrected in a simple manner. The punch is set to the position at which the incorrect hole is located, and the incorrectly punched card is placed as if for punching. A piece of paper card stock, of ordinary index-card thickness, is placed over the plastic card and the perforating die is blocked by placing a suitable piece of metal shim stock under the card. The lever is then depressed, punching the paper card and forcing the paper punching into the hole in the plastic card. No cementing appears necessary; the fibrous insert spreads sufficiently to lock itself firmly in place.

The index designed for the field of instrumentation includes approximately 1000 primary terms (i.e., terms represented by cards). These terms are typed on the Peek-a-boo cards with a preceding number to identify the category (see Figure 6-4). The cards are filed by category and by alphabet within categories. Information to be punched is furnished to the operator in the form of 3 x 5 inch abstract or citation cards, on the reverse of which are the index terms assigned by the document analyst. These "master" cards are serially numbered and are presented to the operator in serial order. Assuming the operator has just completed punching for a prior document, the punch is adjusted to the next position by one rotation of the "units" knob. Each pertinent Peek-a-boo card, selected from the file in turn as called for by the master card, is inserted into the punch, punched and returned to the file. When all the index terms have been thus recorded, the punch is set to readiness for the next document by rotating the indexing knob. When a column of 100 hole positions has been completed (100 documents entered), the fence carriage is returned to the units "zero" position by releasing the split nut on the "units" screw. The "hundreds" screw is then rotated one turn advancing the carriage to the next column, and punching is resumed, document-by-document.

Experience thus far demonstrates a punching rate of 2000 holes per day (approximately 200 documents) for a single machine. This rate allocates an acceptable fraction of total cost to the punching operation. Enhancement of the rate is expected to result from a modified arrangement of the Peek-a-boo file which will stagger the cards in the file to make the headings directly visible, as is done for example with Ekaha cards¹⁶.

When all of the hole locations on a set of cards have been utilized (i.e., 18,000 documents) it is necessary to establish a new set. For the rate of accumulation anticipated to cover the literature of instrumentation this represents roughly one set per year. For the punching operation only one set is active at any one time so that the addition of sets has no influence on this operation. The effect on the searching process is discussed below.

Searching. Since the card bears no imprinted grid from which the position of holes may be read, an overlay, consisting of a transparent sheet ruled with a millimeter grid with appropriately numbered scales, is used. Readout may be accomplished by holding the stacked cards, together with the overlay, in proper register against a bright background. The overlay superimposed on cards to be read, stacked together on a transparent plastic rack, is shown in Figure 6-6.

The search for information starts with a "translation" of the search question in terms of the index vocabulary. The selected terms may be grouped in various combinations of different specificities as a guide to the searcher. This is particularly desirable if the actual search is to be performed by a

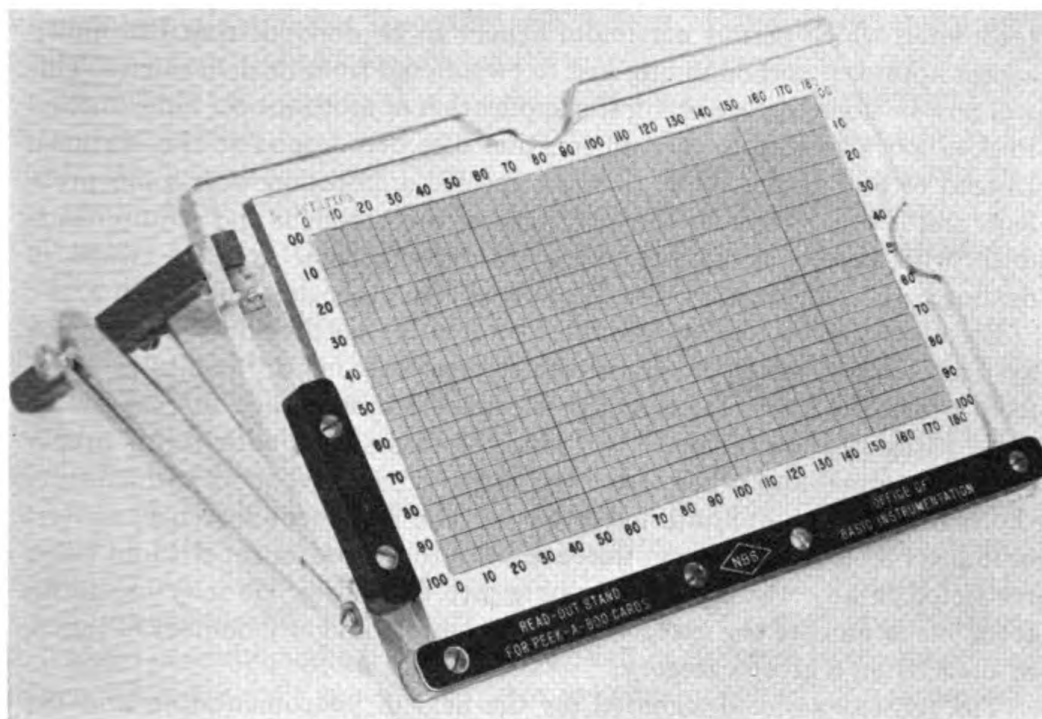


Figure 6-6. Read-out stand for Peek-a-boo cards.

clerk. Instructions to the searcher may request, for example, that restrictive terms be added in a given order until the number of documents disclosed has been suitably limited. The searcher withdraws the requisite index cards from the file, places them on the rack together with the overlay, and notes the document numbers indicated by the intersection of vertical (hundreds) and horizontal (unit) grid lines. It has been found convenient to have the holes fall within squares and to establish the document number by taking first the three digits of the line to the left of the hole and then the two digits of the line below the hole. Entering the "master" file with these document numbers, the searcher obtains the corresponding citations and abstracts.

Refiling may be facilitated and errors in filing quickly found and corrected if a diagonal stripe or groove is marked on the upper surface of the collection of cards as seen in the file drawer.

Application to Instrumentation Literature

In the preceding pages we have discussed an indexing approach based, at least in part, on a search technique which requires that a number of general classes be combined to produce a more specific class. An effective mechanism for carrying out this operation has also been described. It remains to discuss the procedure for setting up such an index and to describe some of the

techniques which permit maximum benefit to be derived from the multi-aspect approach and minimum loss to be suffered from its deficiencies. This will now be done by describing the application of multi-aspect indexing and Peek-a-boo searching to instrumentation files maintained at the National Bureau of Standards. While this is a specific application with some problems and features peculiar to itself, most of the problems and requirements dealt with in setting up the index are quite general and the ways of dealing with them are immediately applicable in other documentation areas.

Broadly, the vocabulary established for the instrumentation indexing system is based on ten categories which represent major "points of view" from which the information seeker is expected to approach the search. To these categories are assigned two kinds of terms, the index terms proper (primary terms) and other terms (synonym or referred terms) which refer the searcher to appropriate index terms. The document searcher, analyst or indexer assigns terms from the categories to the document without being restricted in any way by the grouping into categories. He may use terms from one or more of the categories as appropriate and any number of terms, or none, from a given category.

The ten categories designated for the field of instrumentation and the approximate number of primary terms in each, are:

(1) Measurands (100 primary terms). These are the terms representing the variable, quality, quantity, or condition, the measurement of which is the subject of the document. Typical terms are: acceleration, force, displacement, brightness, magnetism.

(2) Principles (300 primary terms). These are the terms which denote the operating principles of the instruments or methods by which the measurement is performed. Examples are electromagnetic, sonic, radiation, spectra, resonance.

(3) Object (100 primary terms). The materials or things of which the measurand is a property, for example, copper, structure, fluid, celestial body, vehicle.

(4) Instrument Name. This category contains no primary index terms, only referred terms, because no way has been found to generalize the large number of special names without redundancy with other categories. For example, if the measurand is voltage, the instrument is likely to be a voltmeter, if its operating principle is electrostatic the instrument is more specifically an electrostatic voltmeter. The very specific individual instrument names, particularly trade names of instruments, are listed with definitions which reveal the measurand and operating principle, thus referring the searcher to the appropriate index terms in other categories.

(5) Field of Application (60 primary terms). These terms indicate the area of science or technology in which the document originated. This is

frequently quite different from the field defined by the measurand itself and offers a powerful tool for refining the search. For example, if one is concerned with the measurement of acceleration, the fact that a paper on this subject originated in connection with aviation may have a great bearing on the likelihood that the information contained therein is appropriate to the needs of the search. Examples of other fields typically represented in the index are: instrumentation itself, research, medicine, navigation, process industry.

(6) Development Stage (10 primary terms). In the course of development of an instrument various stages are recognizable and papers appear which deal in major aspect with particular stages. Examples of the index terms to be found in this category are: design, construction, prototype, production, testing, calibration.

(7) Function in Instrumentation Process (50 primary terms). The process of measurement is conceived as consisting of a series of functions. Examples are detection, amplification, transmission, display, recording, computation, control. These functions may form a sequence of operations in a given case although no fixed sequence can be assigned to the functions in general. Frequently specific functions are a major subject of a document which can be recognized by index terms drawn from this category.

(8) Character of Document (50 primary terms). Such descriptive characteristics as language, government report, bibliography, book, history, review, are indicated here. Like category (5) this category is quite generally applicable.

(9) Characteristics of Device (50 primary terms). Here are indexed such performance parameters as accuracy, precision, ranges, environmental limitations and such features as portability, etc.

(10) Limitations on Measurand (200 primary terms). This category is designed to cooperate with category (1) to specify the measurand further. Many of the index terms in this category are duplicates of terms in (1). The fact that they are in category 10 gives them a different significance. The word "temperature" in category (1) signifies that the document deals with the measurement of temperature. The same term in category (10) signifies that the measurement in question is somehow limited by temperature considerations; for example it may be concerned with measurement of pressure at high temperatures.

Categorization is expected to provide a number of advantages. First among these is guidance of the thoughts of the indexer and the searcher along similar tracks. Related to this is the added assurance that the indexer (and the searcher) will give consideration to all of the pertinent aspects of documents relating to instrumentation, although it is not expected that every document will have a term from every category. The categories also

serve to distinguish between identical words used with different sense, as in the example cited of the significance of the word "temperature" as a measurand in category (1) and as a limitation in category (10). The use of categories is a recognition of the fact that it is possible, in any information area, to select at least a few groupings of related terms which are universally acceptable and may be expected to remain stable. In making use of relationship of the categories to the terms subsumed within the category, advantage is being taken of desirable features of conventional indexing.

The Index Terms. The indexing terms, conveniently referred to as primary terms to distinguish them from the much larger number of secondary terms (which are cross referenced to the primaries and which are not themselves separately represented by Peek-a-boo cards), at present number about 1000 for the instrumentation reference file. A conscious effort has been made to keep the number of primary terms small. No limit is placed on the number of secondary terms listed in the search dictionary, approximately 2000 at present. It is believed that these should include every term which may be thought of by the technical searcher and should disclose how each is represented in the index.

A number of important advantages derive from a small vocabulary of primary index terms:

(1) A small vocabulary implies a corresponding reduction in the severity of the synonymous-expression problem. The number of meaningful combinations of terms, and hence the number of synonymous expressions of a class (i.e., different combinations of terms which express the same idea), increases rapidly with the size of the vocabulary. Each of these offers the indexer or searcher an alternative choice of combination terms, only one of which is correct.

(2) The labor of finding cards in the alphabetical file, for punching and searching, is roughly proportional to the number of cards.

(3) A small vocabulary will be learned rapidly by the indexer resulting in more rapid assignments and in fewer erroneous assignments. The same consideration holds true for searching.

A small vocabulary is achieved at the expense of a larger dependence on combining terms to express specific index ideas. This increases the number of terms assigned to the average document, thus increasing the probability of false associations and the punching load. A proper balance must be achieved among these opposing factors. In order to avoid reintroducing the word-order problem, composite terms are used very sparingly. In principle, apart from the word-order problem, the accidental nonexistence of a single term to express a desirable index idea is not permitted to exclude the use of that idea as a primary term, since many composite terms have acquired a preferred word order and hence perform in all respects as single terms. It is

frequently found in devising a set of primary terms, that no word is available to express correctly the content of a desired primary index class. In such cases it is useful to select as the index term a word which describes a part of the class and to use this term arbitrarily as a code which represents the class. The meaning of the term then is a mnemonic rather than a precise indicator of subject class content. Thus the word "capacity" might designate a class which also includes "capacitor," "capacitance," and "volume," as well as "elastance," the reciprocal of capacitance. It may, in fact, be useful at times to go farther in regarding the term as a "pronounceable code" and group items in a class for no logical reason other than usefulness in searching. As a result of this kind of technique some of the referrals from secondary terms to primary terms do violence to the accepted meanings of terms and may cause the purist some discomfort.

The lists of primary terms have undergone several major revisions and spot revisions are expected always to be in order.

Document Analysis. The document analyst assigns to a document terms selected from the index vocabulary and he may designate portions of the text which are to constitute an abstract. The term assignments, the reference citation, and the abstract, if used, are noted on a 3 x 5 inch card which becomes the master card for the document. This card is assigned a serial number which designates the location at which its representative hole is to be punched in the Peek-a-boo file. Indexing words which suggest themselves to the analyst, and are not in the dictionary of terms, are noted and at a later date are entered into the dictionary, and assigned "see" references. The all-embracing nature of the dictionary built up in this way is believed to be a vital element of the reference system.

Document analysts are technically trained in fields basic to instrumentation. Experience so far does not suggest that the new approach is at all different from conventional indexing in the need for subject experts to analyze documents or in the level of technical competence required.

Further Developments

The versatility of the indexing system described permits a variety of techniques which have not yet been applied in the instrumentation reference file but which are expected to prove valuable in the future or in other applications. Some of these will be discussed here.

Unconventional Subject Terms. Many concepts, not usually regarded as subject descriptors, provide useful search indicators. These can often be treated in the same way as normal index terms. For example, it is usual to regard the author index as quite separate from the subject index. Nevertheless, reference questions frequently arise which require interaction of author and subject. It is also a matter of considerable convenience to be able to maintain one general file instead of a number of special-purpose files.

Authors can be indexed in Peek-a-boo in the same way as subjects. To avoid inefficient use of cards and a great increase of the number of cards required, each author is not represented by a card but is alphabetically grouped with other authors. For example, the first three consonants in the name might be used to define groups requiring 3 sets of 26 alphabet cards for their identification. Interaction between author groups and very general subject classes is used to separate a particular author from others in his group.

Chronology, "in-file" indications, journal titles, institution of origin, and country of origin may also be the basis for search cards. Numerical information may be recorded on Peek-a-boo cards by using cards to designate digits to any degree of fineness. Other cards may be used to record exponents of 10. A precision of 1 per cent (2 significant figures), and a range of 20 orders of magnitude, may be recorded with 20 digit cards (2 sets of 10), 10 exponent cards, and 2 cards to identify sign of the exponent. Thus the number 1200 may be recorded by punching a "1" card, a "0.2" card, an exponent "3" card, and a "+" card; and the number 0.12 by punching the same digit cards, an exponent "1" card, and a "—" card. Various other systems can be devised to exchange economy in total number of cards for economy of punching.

Converse Searching. It is sometimes desirable to be able to search for the absence of a given characteristic. That is, in contrast to the normal search, one wishes to find all documents (or other entities) which have *none* of a group of index characteristics. Such a situation occurs, for example, in an application of the Peek-a-boo principles to indexing infrared spectra being developed by Savitzky²⁷. In searching infrared spectra for identification of constituents of an unknown mixture of which the spectrum has been measured²⁸, it is desired to find all compounds which have no infrared absorption bands in any of a number of frequency ranges. It is, of course, possible to have "no-band" cards for this purpose. This would, however, require that each "no-band" card be punched in the large number of cases in which no absorption band is present, imposing a very large punching burden. It is possible to reduce greatly the burden by inverting the usual punching operation. In this inversion, a transparent card, or a totally pre-punched card, is used instead of the usual opaque card. Then it is not necessary to punch the multitude of cards which do *not* characterize the document (or spectrum), only to treat the cards which *do* characterize it, by filling the perforation (on a pre-punched card) or otherwise obscuring a spot (on a transparent card). To minimize punching, each index card

²⁷ Savitzky, Abraham, Perkin-Elmer Corp. Private communication.

²⁸ Baker, A. W., Wright, N., Opler, A., Automatic Infrared Punched-Card Identification of Mixtures. *Anal. Chem.*, 25, 1457 (1953).

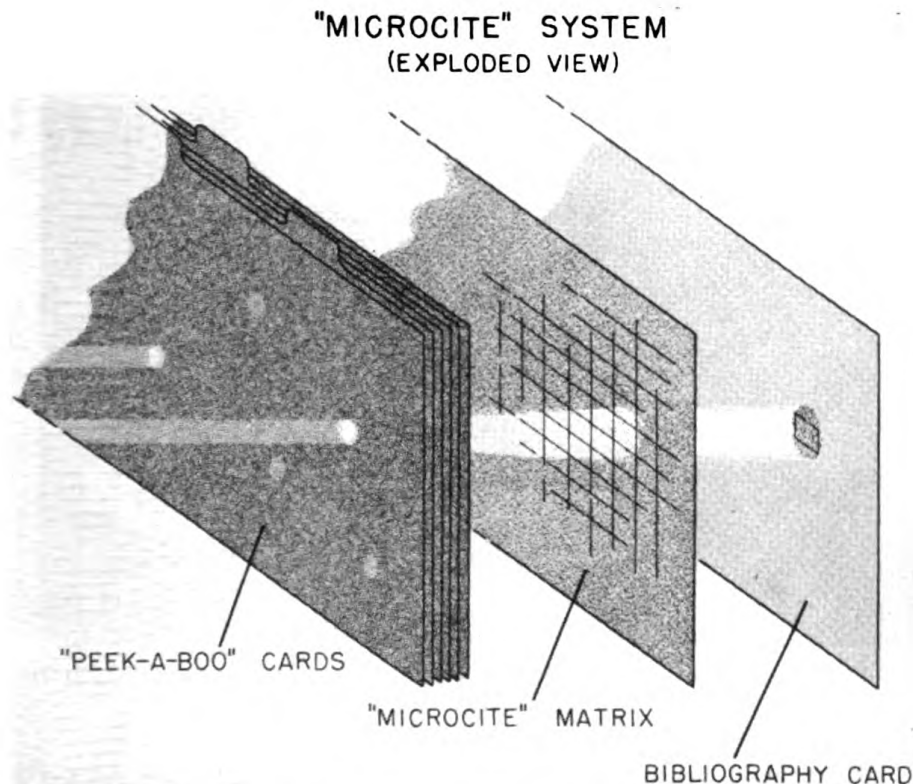


Figure 6-7. "Microcite" system, principle of operation.

could be paired with its pre-punched converse and the operation of punching could simultaneously fill the corresponding converse hole (as in the technique of repairing errors previously described). Photographic techniques for preparing positive and negative copies are also obviously applicable.

"Microcite". In the section dealing with the comparison of subject cards with document cards, it was pointed out that the serial numbers yielded by the subject-card search constitutes an intermediate step which carries with it some undesirable features. To overcome this shortcoming a development has been undertaken, which has been named "Microcite," the principle of which is illustrated in Figure 6-7. With the Peek-a-boo search set is associated a matrix film on which is photographed, with suitable reduction, the equivalent of the master card (3" x 5" card) used in normal Peek-a-boo searching (i.e., citation, title, abstract, etc.). Each of these is located on the matrix film in the position dedicated to the corresponding document on the punch cards or in a related position. All of the area of the film can be utilized for the photographs because, in contrast to perforations, the photographed areas require no supporting areas. In using Microcite to prepare bibliographies in answer to reference questions, the film matrix (a negative in this case) is to be sandwiched between the Peek-a-boo search stock and a

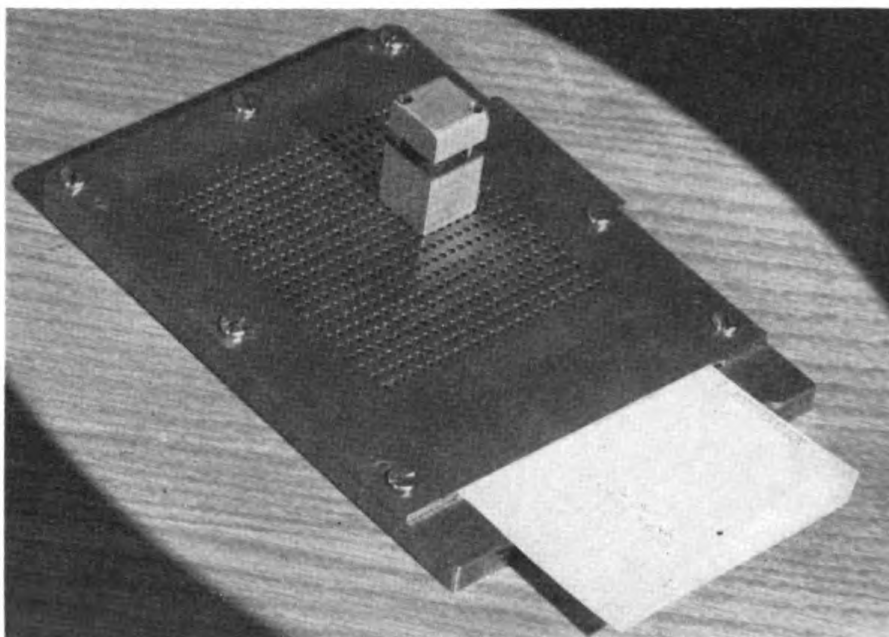


Figure 6-8. Card punch for 1000 document cards used in experimental Microcite system.

suitable photographic printing paper. Light passing through the Peek-a-boo holes provides illumination for printing, thus printing only the selected references. A suitable diffuser between the Peek-a-boo stack and the film matrix permits the small hole to illuminate the larger area of the master card image.

An experimental version is based on a card with 1000 hole positions. Photographic reductions of 30 to 1 are used and a typewritten area of approximately 3 x 5 inches is recorded in the matrix area devoted to each hole. The simple punch developed for this experiment is shown in Figure 6-8. The kind of bibliography card obtainable with the experimental microcite is illustrated in concept in Figure 6-9.

Extension of Microcite to the main Peek-a-boo instrumentation reference collection will require 12 film matrices (or one larger matrix) for each Peek-a-boo set if no loss of master-card information or higher reduction ratios are assumed. With the very small hole areas involved, further development beyond the experimental version will be necessary to accomplish illumination of the appropriate film-matrix area for photographic printing of bibliographies. However, another use of Microcite, possibly more important than the preparation of microprint bibliographies, involves no such development problems. This is the use of Microcite in guiding the search process by providing for observation of the course of the search. For this purpose the Peek-a-boo holes are used only to locate the corresponding microphoto-

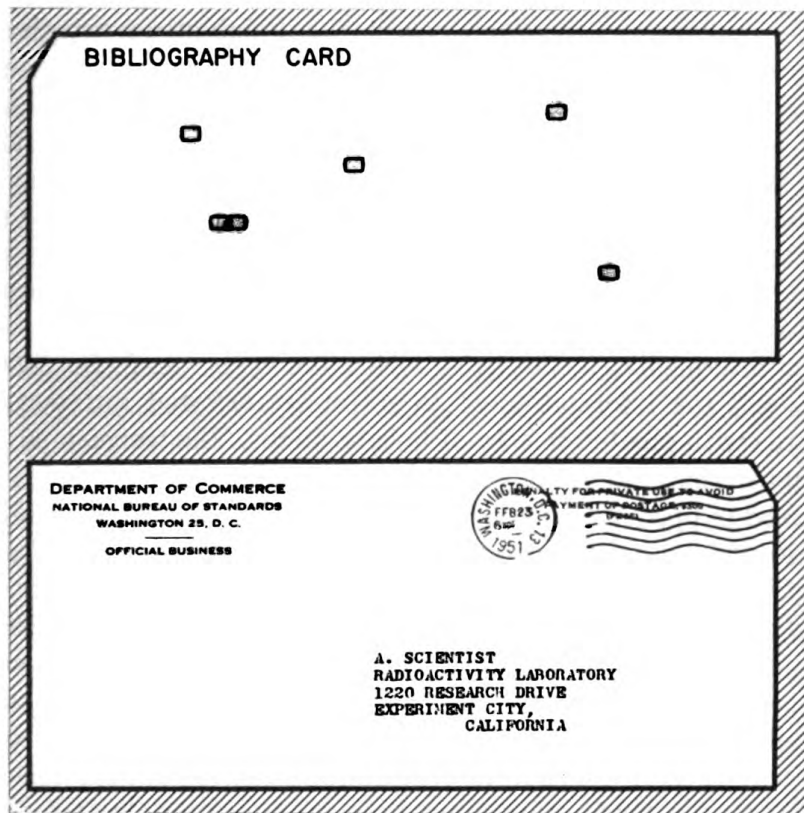


Figure 6-9. Concept of bibliography card product of Microcite search.

graphed citation on the matrix and not to illuminate the area for reading. Reading is accomplished by reflected light, with a suitable magnifier, microscope or microprint reader or by projection. Construction of suitable equipment for this application is underway for preparation of film matrices for the main instrumentation file.

Application to Large Collections. The use of Peek-a-boo cards for large collections involves either multiple sets or larger cards. The use of multiple sets is quite straightforward and appears to be feasible at least for a small number of sets. Experience does not as yet permit a limiting number to be estimated. Great increases in card size appear also to be feasible. For example, a card approximately 42 x 22 inches, accommodating 500,000 documents, does not appear to be unreasonable for manipulation in searching or for the demands made on dimensional stability of cards. Manual punching would probably be abandoned in using such cards. Other reasons exist for giving serious thought to further mechanization of the punching operation. Punching as it is now performed is a fairly costly operation (though minor as compared with document analysis), and adds complexity to production scheduling.

A revised operating procedure, now in the design stage, promises to improve punching efficiency and to permit extension to much larger cards. This procedure is based on the use of preliminary Peek-a-boo sets, to be punched directly by the document analyst. This will replace handwritten notation by the analyst and should not consume additional time. In any event, the analyst's work is paced by the thinking necessary for effective analysis and not by the writing or other mechanical operation he needs to perform. An intermediate card of 1000-document capacity appears suitable and it is planned to use the Microcite punch (Figure 6-8) for this purpose. The intermediate sets of the various analysts are to be collected at intervals and the information of these sets transferred automatically to larger cards with greater density of hole positions. The transfer punch used to accomplish this is still to be constructed. A number of advantages are seen in this *modus operandi* beyond those which initiated its consideration. It lends itself very well, for example, to a "contributing analyst" organization of effort. The punch is simple and efficient in operation; intermediate card sets can be imprinted with index headings by automatic machinery; the individual sets, identified by assigned blocks of numbers, become sub-collections of special utility to the contributor; the sets can be combined in a variety of ways with other sets, not necessarily the entire collection, for special purposes; the subsets provide a means for production of replicate main sets for distribution or publication.

Very Large Collections. It is not possible at this point to guess at the maximum size of collection for which the visual Peek-a-boo technique is effective, or even to predict whether the Peek-a-boo technique or the multi-aspect classification itself will be first to break down with increasing size. It is useful, nevertheless, to consider the characteristics of electronic systems analogous to Peek-a-boo to see whether these may be expected to handle very large collections which are becoming of more and more common concern. The basic advantages of subject card *versus* document card are certainly not limited to the punched-card medium. The disadvantages of subject cards are in fact less significant when the concept is extended to more automatic operation. There would certainly be less occasion to distribute very large systems than smaller ones and increasing document capacity implies only an increase in size and not a less convenient operation. A number of electronic analogs of Peek-a-boo have been outlined of which the following is an example.

On a multitude of magnetic-recording channels (magnetic wire, tape or drums may be employed for example) are to be recorded "bits" or signals (pulses) whose positions along the channels represent document serial numbers. The individual channels themselves then are analogous to Peek-a-boo cards and represent index terms. The serial-number "pulses" might

be recorded as amplitude modulations of a carrier of suitable frequency, with the carrier-wave count itself used to identify position along the wires. Alternatively, spots on a paper tape could also represent the "bits." Terms assigned to a document would be inserted by a typewriter keyboard positioning a recording head to the appropriate channel. Reading out would be accomplished with a similar keyboard selection of the index term followed by a search for coincident bit signals or pulses along the several channels selected for search. Carrier signals stored on the channels might serve to monitor the reading speed and, via controlled delays, to compensate for minor phase shifts to make extremely precise synchronism of the channels unnecessary. The result of the search might be preserved in an intermediate storage whence it would be used to actuate devices to deliver full-size copies, abstracts, or mixtures of the two as instructed by the searcher, or a sample of the results could be displayed to guide adjustment of the search terms before documents disclosed by the search were delivered.

Conclusion

In this chapter, the mechanics of an information search system have been described in some detail. Further details, including drawings for constructing some of the devices, can be made available upon request to the authors. A number of expressions of interest in the commercial manufacture of equipment for peek-a-boo searching have been received. Most of the devices concerned do not appear to be patentable and any patents which do result from the work described here will be assigned to the U. S. Government and will therefore be available on a royalty-free-license basis. The availability in this country of simple inexpensive equipment would provide an important stimulus to application of Peek-a-boo in fields for which it appears suitable.

Extension to Other Fields. The Peek-a-boo principle is versatile enough to make it suitable for very small collections as well as for moderately large ones. The simplicity of material and equipment enables one to establish a file with minimum investment. A variety of inexpensive cards in many sizes can be adapted to Peek-a-boo in combination with simple hand punches. This simplicity of equipment, together with the great rapidity of searching, suggests usefulness in such fields as police files, medical diagnosis, law, catalog and inventory files, real estate, personnel selection, and many others. The open endedness with respect to subject which is characteristic of Peek-a-boo, minimizes the losses resulting from errors in setting up the index. There is a need, nonetheless, for promulgation of basic index lists in a variety of fields which would enable many users to start Peek-a-boo files with a minimum of lost motion.

Chapter 7

A UNITERM SYSTEM FOR REPORTS

Charles E. Zerwekh, Jr.

*Research and Development Division,
Humble Oil & Refining Company, Baytown, Texas*

Introduction

Reports generated within an industrial concern constitute one of the principal sources of information and know-how which can serve as the basis for future research and development activities of the organization. Since much of the subject content of such reports never gets published, industrial organizations are confronted by the problem of indexing or classifying their own reports.

The impetus to undertake application of the Uniterm system to the reports generated by the Research and Development Division of the Humble Oil & Refining Company was provided by the rather glaring inadequacies of previously used indexing systems. One of these systems was based on title indexing. This index was maintained on 3 x 5 inch cards on which, together with the index entries, were recorded the usual bibliographic data, such as the name of the issuing organization, the date of appearance of the report, etc. It was observed—as is so often the case with titles—that this form of indexing was often misleading or inappropriate as far as indication of important aspects of the subject content of reports was concerned. Also, the frequent inversion of entries as exemplified by “Cracking, catalytic” was a source of delay and exasperation in using this index. The inadequacies of this system led to attempts at more systematic indexing, the purpose being to analyze the subject content of the reports more carefully. Here again 3 x 5 inch cards were used for each subject entry set up, and on each card were recorded, in addition to the usual bibliographic data, all other subject headings under which the report had been indexed. The reports were not indexed in sufficient detail to provide the desired measure of accessibility to technical information. At the same time, even with a limited amount of indexing—an average of five cards per report—the typing problem was severe. The number of cards required per report was too small to justify mimeographing, but the typing of multiple copies of file cards, on the other hand, was time consuming and burdensome. As a consequence there were frequently intolerable delays between the date of

issuance of a report and its indexing. Therefore, the Uniterm system¹ was adopted to provide faster and more adequate processing of reports.

Basic Procedures of the Uniterm System

The operation of the Uniterm system is simple and easily described. A separate card is set up for each of the terms used to designate the important aspects of reports. On the same card are listed the identifying numbers of all reports for which the term in question is an appropriate subject designation. The report numbers on each card are arranged in the following manner for convenience in conducting subsequent searching operations: The card is divided into zones as shown in Figure 7-1. The numbers of various reports are then arranged in these zones, depending upon the final digit in the number. Furthermore, within each of the columns, as Figure 7-2 shows, the report numbers are arranged in ascending numerical order. Before describing the use of a Uniterm file, let us consider how it is established.

The analysis of the subject content of reports by the Uniterm system consists of the following steps. The serially numbered report is first read and a decision is made as to which terms are appropriate to designate as important aspects of its subject content. This review of the report, the decisions as to important aspects of subject content, and their expression by appropriate symbolisms—in our case by words—has much in common with the performance of alphabetized indexing. The next step is to locate in the file those Uniterm cards which correspond with the words that have been selected for designating important aspects of subject content. If by chance a Uniterm card has not been provided for any one of the words so selected, a decision must be made as to one of two possible courses of procedure.

Absorber; Absorbent; Absorption									
0	1	2	3	4	5	6	7	8	9
40	1	2	3	4974	4695	146	1157	708	1239
5150	41	262	123	5024		2306	3737	868	1459
	251	602	3783	5064		2376	4107	1548	1549
	4971	1042	4363	6764		4796		4908	3719
		1542	4913	6774		4906		5338	6509
		4262	5413			4916			
		5412	5503			5206			

Figure 7-1. Uniterm card divided into zones (numbered 0-9) in which are entered serial numbers of documents which contain the key words or ideas as represented by the card heading: "absorber; absorbent; absorption." The terminal digits of the document numbers determine the zone in which the number is entered.

¹ The "Uniterm system" of coordinate indexing was developed by Dr. Mortimer Taube and his staff at Documentation, Inc., Washington, D. C.

Absorber; Absorbent; Absorption

0	1	2	3	4	5	6	7	8	9
40	1	2	3	4974	4695	146	1157	708	1239
5150	41	262	123	5024		2306	3737	868	1459
	251	602	3783	5064		2376	4107	1548	1549
	4971	1042	4363	6764		4796		4908	3719
		1542	4913	6774		4906		5338	6509
		4262	5413			4916			
		5412	5503			5206			

Spectra

0	1	2	3	4	5	6	7	8	9
120	1951	1432	5713	1064	2885	4576	1277	1278	1279
190	2451	2567	6263	2894	6025	6026	2877	5728	4489
	2471	4502		4524		6106	4107		5069
	6121	5662		5714		6506	4697		
	6421	6262				6636	5707		
							6247		
							6457		
							6517		

Infrared

0	1	2	3	4	5	6	7	8	9
80	71	72	443	1064	1065	66	77	108	1249
1250	241	112	4173	2894	2885	1896	1897	118	1459
1650	4091	1432	5413	4144	4165	4086	2017	4088	3079
5530	4701	3132	5723	5344	4725	5266	2487	4138	3539
6480	5661	4652	6273	5724	5295	5296	4107	4388	4119
		4712		6764	5375		4217	5278	4139
		4942					5267		4199
		5322					5297		4489
		5662					5537		5359
		6452					6457		5719
									6389
									6449
									6599

Figure 7-2. A search is conducted for all documents in the Uniterm file which contained the key words or ideas: "infrared," "spectra," and "absorption." The three cards—as shown—are selected, and are compared to determine which document numbers are in common to all three cards: viz., number 4107.

It may be decided that, for the term in question, a synonym or near-synonym for which a Uniterm card has already been provided may serve to designate the subject content aspect. In this case the report number will be entered on the existing Uniterm card selected as being appropriate. Alternatively, it may be decided to establish a new Uniterm card. This question when it arises is an important one, and will be discussed further. After the Uniterm cards have been located in the file or new cards provided if necessary, the report number in question is entered on the various cards following the procedures described and illustrated in Figures 7-1 and 7-2.

The Uniterm file is used for searching and retrieving as follows: If it is desired to identify all the reports relating to a given subject such reports will be listed on a single Uniterm card, provided, of course, the subject in question has been set up as one of the Uniterms. In our system, for example, one will find listed on a single card all reports embraced by the system which relate to the subject "Fungicide." Many research requirements cannot be so simply expressed, however. If a search were directed to a broader subject, such as biologically active materials, it would be necessary to call attention to the entries on two or more Uniterm cards. Thus in this system we might consult reports whose numbers had been entered on the Uniterm cards for "fungicide," "insecticide," "toxicity," and possibly various other cards.

Another possibility is that we may be interested in fungicide activity, only if it involves some one organism or class of organisms. In our file, which is devoted principally to petroleum processing, we have not found it advantageous to index in detail with regard to the organism to which fungicides are applied. The number of reports in the file, however, dealing with fungicides is not large. As a result, a person desiring such information is well served by the Uniterm file if his attention is directed to this small number of reports. If in the future a large number of reports dealing with fungicides were to enter our file, it would not be difficult to reanalyze these reports, and to set up additional Uniterm cards to correspond to major groups of fungus organisms, or individual species of organisms in case they were important. This is an important feature of the Uniterm system, and the possibility of carrying through a reanalysis of certain reports, at the time that such reanalyze becomes necessary, should be kept in mind as an important advantage of the Uniterm system.

The subject contents of most scientific and technical papers usually relate to a multiplicity of things. Similarly requirements for information may often be stated as involving two or more terms. Thus an information requirement might be exemplified as relating to the use of furane chemicals for fungicides. In our system the key terms would be "furane" and "fungicide," both of which appear on Uniterm cards. A list of report numbers which might be of pertinent interest to such an information requirement may be

readily compiled by using this file. This would be done by consulting the cards for the terms "Furane," and "Fungicides," and noting the report numbers that have been listed under both of these terms. There is, of course, the possibility that one or more reports may have mentioned a furane compound and also referred to fungicides, without the furane compound necessarily having been used as a fungicide. Since our system embraces a relatively small number (about 7,500) of reports at the present time, the possibilities of being directed to such reports have not been proved to be a practical problem.

At the Humble Oil & Refining Company the reports themselves are maintained in a classified arrangement and the reports required for a given problem or situation may be located by this system. For reports dealing with a specific subject, it is often more convenient to utilize the Uniterm card file. We maintain an auxiliary card file, arranged numerically by serial number of the reports involved. On each of the serial number cards is recorded the usual bibliographic data concerning the report in question, as well as the class designation which is used for filing the individual reports. Thus a report can be located by first identifying it by its serial number in the Uniterm files and then using this number to locate its position in the classification scheme, and finally locating the report in the classified file.

Observations Made in Applying the Uniterm System

Application of the Uniterm system at the Humble Oil & Refining Company was initiated in the Fall of 1953. It was immediately observed that the reports could be indexed much more rapidly by applying Uniterm procedures than was possible with the subject heading list previously used. A three-month backlog of reports awaiting indexing was cleared up in about two weeks. Furthermore, it was observed that the Uniterm file very markedly accelerated the speed of searching and retrieving material in the file. A quantitative estimate of the time saved is difficult to make, since the depth of indexing was markedly increased when the Uniterm system was installed. As a result it is now possible to locate material that could never have been found in the system used prior to Uniterm. In a brief comparative study, searches of three widely different types were completed in one-half to one-tenth the time required before installing the Uniterm system. At first scientific and technical personnel were a little bewildered by the system, particularly by the fact that the customary bibliographic data was lacking on the Uniterm cards, and that instead the reports were identified exclusively by serial number. A minor amount of explanation, however, sufficed to clear up this difficulty, and it was observed that with a little experience the technical and scientific personnel were able to use the file expeditiously and advantageously. The advantages observed in practical application and the inherent simplicity of the system have led to the conclusion that its capabilities answer our present report indexing problem.

On the basis of experience, we are convinced that the system will continue to function well for at least several years to come. When more than 20,000 reports have been incorporated into the file, it is anticipated that certain difficulties may appear. For example, as more and more reports are embraced by the Uniterm file, an increasing number of reports will have to be listed on the various Uniterm cards. For some of the Uniterms, more than one card will be needed to list all the reports. There will be an increase in the amount of time required to compare lists of reports pertaining to two (or more) Uniterms in searching for reports pertaining both to "Furane" and to "Fungicides." Although such difficulties are foreseeable, it seems unlikely that they will seriously impair the usefulness of the Uniterm file for several years. At the present time, our Uniterm file embraces about 7,500 reports, and it is expanding at the rate of about 500 reports per year. In addition we are indexing into the file reports written and filed before the Uniterm system was started. There are about 5,000 of these reports now awaiting our attention. A considerable number of older reports have already been incorporated in the file.

It might also be pointed out that our file is restricted almost entirely to reports originating with the Humble organization or its immediate affiliates. Also included is a small number of confidential reports originating with contractors and with similar persons. Progress reports are not indexed in Humble's Uniterm system.

As with any system for processing the subject documents for subsequent retrieval, a number of policy decisions had to be made. One of these decisions, as mentioned above, concerned the range of material to be embraced by the system. Other decisions related to the analysis of the subject content of the reports, particularly what aspects of subject content should be taken into account and what Uniterms should be used to designate those aspects that are considered important. It should be emphasized that it is essential to distinguish between these two operations. One of them is the decision as to which aspects of subject content are indeed important. In this connection, quite naturally, we are guided by the range of interests of the Humble Oil & Refining Company. In general, it may be said that the reports which originate in our own research laboratories are indexed in considerable detail while many reports from affiliates are indexed relatively sparsely. For example, at the present time our Company's limited interest in grease and asphalt research would result in reports on these subjects being indexed in relatively little detail. Certainly, however, such reports would be entered on the Uniterm cards for "Grease" and for "Asphalt." If, at some future time, it should prove advisable to index reports on these subjects in more detail, these Uniterm cards would quickly direct our attention to the reports requiring further processing.

As noted above, once a decision has been made as to what aspects of sub-

ject content are of importance, there still remains the important decision as to selection of suitable terms to designate such aspects. At first individual words were used almost exclusively as Uniterms. Thus the number of a report discussing methyl alcohol was entered on one Uniterm card for "Methyl" and on another card for "Alcohol." In practice, this procedure was found to introduce considerable inefficiency into the operation. A large number of entries were found to accumulate rapidly on the "Methyl" card and also on the "Alcohol" card. Furthermore, the need to identify reports relating to methyl alcohol occurred sufficiently often that an excessive amount of time was being devoted to comparing report numbers entered on these two cards. As a result, it was decided to make out a card for "Methyl Alcohol." In this instance, therefore, the term "Methyl Alcohol" is one of our Uniterms. It might be pointed out that redoing the file to establish the card on "Methyl Alcohol" was not difficult. All that was required was to determine which report numbers appeared on both the Uniterm cards for "Methyl" and for "Alcohol" and to list these report numbers on the card set up for "Methyl Alcohol."

It was observed, however, that it was advisable to maintain a generic card for "Alcohol." It was decided further that in the future, we would not enter on this card reports that are concerned with "Methyl Alcohol" only. Rather a cross reference was provided for the "Methyl Alcohol" and "Alcohol" entries in the Uniterm subject list. In this list the "Methyl Alcohol" entry is accompanied by the notation, "Methyl Alcohol," *see also* "Alcohol," and a corresponding entry under "Alcohol" reads "Alcohol," *see also* "Methyl Alcohol." Furthermore, on the "Methyl" Uniterm card we note only those reports in whose subject matter the methyl group itself is an important entity, as exemplified by the methyl radical, $\text{CH}_3\cdot$. Enough has been said, perhaps, to make the point that the Uniterm system must be applied with care and understanding to achieve optimum results.

A similar situation was observed in connection with the Uniterm cards originally established for "Catalytic" and for "Cracking." It quickly became apparent that it would be advantageous to set up a Uniterm card for "Catalytic cracking" for much the same reasons that a card was set up for "Methyl alcohol." However, just as we found it advantageous to maintain an "Alcohol" Uniterm card but to restrict its use to alcohols other than methyl alcohol, we found also it advantageous to maintain a "Cracking" Uniterm card, but to restrict its use to instances not covered by "Catalytic cracking." Corresponding cross-references for "Catalytic cracking" and "Cracking" were set up on our standardized list of Uniterms along the lines outlined above for "Methyl alcohol" and "Alcohol." Our technical men have pointed out a number of similar instances in which it is advisable to provide a Uniterm card for such terms as, for example, "Solvent extraction." In this instance, however, the number of all reports that pertain to

Bromination, see also *halogenation*
Bromine, see also *halogen*
C/H, see *carbon plus hydrogen plus ratio*
Chlorination, see also *halogenation*; *sulfochlorination*
Chlorine, see also *halogen*
Concentrate; *concentration*; *concentrator*
Condensate; *condensation*
Condensed; *condenser*
Digital; *digitization*; *digitizer*
Gage; *gaging*
Gauge, see *gage*
Halogen, see also *bromine*; *chlorine*; *fluorine*; *iodine*
Halogenation, see also *bromination*; *chlorination*; *sulfochlorination*
Liquid; *liquor*
Monobromoxylene, see *halogen plus xylene*
Monomer, see also *monoolefin*; specific monoolefin such as *ethylene*, *propylene*, *butylene*
Paraxylene, see *p-xylene*
PCP, see *pentachlorophenol*
Toxicity; *toxicology*
Versene, see *chelate*

Figure 7-3. Some representative subject headings and cross-referencing procedures.

solvent extraction as well as to other types of extraction are being entered on the "Extraction" card. In addition, we are entering the appropriate report numbers on both "Solvent Extraction" and "Solvent" cards. Thus the "Solvent Extraction" card may be regarded as providing a ready made listing of report numbers that are entered both on the "Solvent" and the "Extraction" cards. This further illustrates the way in which we have carefully controlled the use of terminology to insure the effectiveness of the Uniterm system.

In applying this system, we have found it highly advisable to exert careful control of synonyms in selecting terms to be set up on Uniterm cards. We have drawn up a list of terms for which Uniterm cards have been established, and we have found it to be essential in insuring consistency in the use of terminology, particularly when consulting the Uniterm file to identify reports of pertinent interest to a given problem or situation.

During the initial phases of applying the Uniterm system, there was a marked tendency for different synonyms to be set up on individual Uniterm cards. Such situations may be exemplified by a portion of our list of the terms for which we have Uniterm cards in our files. As shown in Figure 7-3, such closely synonymous terms as *halogen*, *chlorine*, and *sulfochlorination* obviously require reconsideration, if use of the file is not to involve either (1) excessive time in consulting a number of cards to cover a single subject or (2) failure to cover a single subject by overlooking the need to check a number of Uniterm cards. For such synonyms and near-synonyms,

it is necessary, for reliability and efficiency, to provide a single Uniterm card and to be careful to see that reports relating to the subject in question are entered on the card for the term selected.

In Figure 7-3, various sets of synonyms and near-synonyms have been drawn together into adjacent positions by alphabetization of the various words. It must be noted, however, that alphabetizing cannot be relied upon to direct attention to all instances in which synonyms or near-synonyms should be grouped and a single Uniterm card prepared for the group of terms. Thus, for example, we have found it advisable to prepare a single Uniterm card for "Doucil" and "Zeolite" and also a single Uniterm card for "Drainings" and "Sludge." In setting up such cards some one of the synonymous or closely-synonymous terms is selected for the Uniterm card, and the synonyms and near-synonyms are cross-referenced in the standardized listing of terms, as shown in Figure 7-3. The advantages of providing "*See also*" references have been discussed previously in connection with setting up Uniterm cards for "Methyl alcohol," "Catalytic cracking," and "Solvent extraction." It is perhaps obvious that one may learn a great deal from analogous procedures as practiced in subject indexing. We have found it highly advantageous to set up our Uniterm cards for chemical compounds on the basis of the well-established nomenclature practices of "Chemical Abstracts." Thus the name of a compound which consists of two or more words, e.g., methyl alcohol, is regarded as a single Uniterm and such terms are not split into component words when establishing Uniterm cards. Occasionally, we find it helpful to our scientific and technical personnel to take into account the abbreviations and chemical slang that they use habitually. Thus we may use "MPK" as a term to denote methyl-propyl ketone.

At the present time, the terms which are included in our lists and for which Uniterm cards have been prepared number about 3000. It seems likely that this list will increase rather slowly in the future, with most of the foreseeable expansion being devoted to cross-reference entries of the "*see*" or "*see also*" types.

Conclusion

Our experience with the Uniterm system might be summarized as follows: We have found that it simplifies and expedites our indexing operations. Furthermore, accessibility to the information in our reports has been considerably improved. However, in applying the Uniterm system, it is highly advantageous, if indeed not necessary, to insure that terminology is used in a consistent fashion. Careless use of terminology may result in reports of pertinent interest to a given subject, problem, or project being scattered over a range of Uniterm cards, with the possibility that such material may be overlooked when making a study.

Chapter 8

AN IMPROVED ANESTHESIA RECORD CARD*

MAX S. SADOVE, M.D. AND MYRON J. LEVIN, M.D.

University of Illinois, Chicago, Ill.

AND

Veterans Administration Hospital, Hines, Ill.

The search for a simple, satisfactory flexible system of maintaining anesthesia records has been going on for at least a decade. Each newly devised system has been an improvement over the previous ones, only to become itself obsolescent as new anesthetic agents and techniques are introduced. Certain basic requirements of the joint council on Hospital Accreditation and other official bodies must be met. A complete anesthesia record must be filed within the patient's clinical chart. The clinical records should be able to serve as aids in teaching and learning and they should also be valid as medico-legal documents. Each record must be specific for the job it performs; but in addition, it must permit future expansion as the particular field it serves expands.

Anesthesia records should be so simple that a competent secretary or a medical student can complete them and understand them. This is especially necessary because even the trained anesthesiologist cannot give his fullest attention to his patient while completing a complicated record.

Another essential is that information be recorded in such a manner that it will be available for future use in compiling statistics as well as in studying a given case.

Finally, in a given system of anesthesia records, it is important that space be available for all current technics, agents, methods and all data about the patient such as his physical condition, pre- or postoperative complicating factors, cause of death, etc. It is important that provision be made in the chart for everything that may possibly occur during the conduct of a case. In the previously existing anesthesia records the gathering of statistical data regarding preoperative and postoperative complications has been particularly difficult. Newly developed technics such as the use of hypo-

* Acknowledgment is due to Mr. Erwin L. Vaughn of E-Z Sort Systems, Ltd., for his major part in the development of the Illinois Anesthesia Punch Card.

thermia, hypotension or mechanical respiration-assisting or controlling devices which have come to be accepted standard anesthetic technics, cannot be easily coded on the preexisting forms.

With the above criteria in mind, a new anesthesia record form was devised, which we believe fulfills most of the requirements.

An edge punched card could, if properly designed, fulfill all of the criteria for anesthesia records listed above. Several such cards have been used successfully for the past few years in various institutions throughout the country. The Chicago Keysort card was taken as a basis from which the present card was developed. At a casual glance this card appears exceedingly complicated. Yet on closer inspection, the seeming complexity resolves into its most important advantage. This is the fact that a minimum of writing is necessary to complete the required entries. Almost every conceivable presently used technic, agent, and complication is printed on the card, requiring only a minimal amount of checking. Once the user is familiar with the card it becomes very simple to maintain records with it. The card is so constructed that the front and back sides are each separate units, making it unnecessary to turn the card during the conduct of an anesthetic procedure. The front will be discussed first and then the reverse side.

The front or face of the card contains the following data (see Figure 8-1):

Name
Physical Status
Age
Sex
Ward
Room
Register Number
Anesthetic Number
Date
Proposed Operation
Preoperative Diagnosis
Pre-Anesthetic Medication
Identity
Positive Findings
Preparatory Therapy
Graphic chart with code for Graphing
Remarks column for items not coded or listed
Agents: Primary and Secondary for Induction, Maintenance and Emergence
Methods: Induction, Maintenance and Emergence
Types of Airway Maintenance
Surgeons
Anesthetists
Postoperative Diagnosis
Operations Performed
Left-hand margin (punched)
Sex
Special Interest

Anesthesia Time
 Operation Time
 Special Studies
 Anesthetist
 Preparatory Therapy
 Physical Status
 Age of Patient
 Year
 Right-hand margin (punched)
 Site of Operation
 Position of Patient
 Level or Plane
 Classification of Operation
 Death Analysis
 Sections "B" & "C"

The front of the new anesthesia record card is 10½ inches in width and 8 inches in height. The central portion of the card is a graphic chart similar to those used by most anesthesiologists, but with several important advantages over most of them. These are as follows:

1. The graph extends over a 4½-hour period whereas most of the previous cards do not exceed 2½ to 3 hours. This permits the use of a single card for most operations, whereas previous cards quite frequently required a second sheet. If the operation is shorter than 2 hours nothing is lost since

The form is a detailed anesthesia record card. It features a large central grid for recording data over a 4.5-hour period. The grid is organized into columns for time (0 to 4.5 hours) and rows for various parameters. The parameters include heart rate, blood pressure, oxygen saturation, and anesthesia level. The form also includes sections for patient information, anesthesia details, and a right-hand margin with punched holes for sorting. The grid is divided into columns for time (0 to 4.5 hours) and rows for various parameters like heart rate, blood pressure, oxygen saturation, and anesthesia level. The form is surrounded by a border of punched holes.

Figure 8-1. Illinois E-Z sort anesthesia punched card (front).

a record is required even for a 10 minute procedure. The increasing frequency of lengthy operative procedures makes this feature of the card financially attractive as well as convenient for the user. It also conserves filing space.

2. Six *blank* lines are provided for the designation of *Agents*. This was done to save space, since all the agents could not be listed by name. It offers the additional advantage in that the primary or most important anesthetic agents can be listed first with supplemental agents given on the lines below or, if desired, these may be listed chronologically with the first one used on the top line and each succeeding one on the line below. Even in these days of polypharmacy it is rare that more than four or five agents would be employed in the conduct of any single anesthetic, so that the use of six lines leaves ample room for even the most extreme cases.

3. Two separate lines are provided for the graphic representation of muscle relaxant drugs since these are now extremely important in the conduct of anesthesia, but are considered to be adjunct drugs rather than anesthetic agents *per se*. A box at the end of each line is provided for recording the total dose of these drugs.

4. Two lines are available for recording intravenous therapy—saline, dextrose, blood, etc. These are to be used in the manner described by other authors, but reiterated here:

— represents blood volume expanders and plasma

~ represents watery solutions such as dextrose or physiologic saline.

— represents whole blood.

In each instance, the addition of a new bottle of solution is marked with a vertical arrow, thus '↑' and the kind and quantity of solution is written

18 ga 5% D/W NSS
in. For example: RA ↑ 1000 ↑ 500 would indicate that 1000

ml of 5% dextrose in water was started via an 18-ga. needle in the *right arm* at the first vertical arrow, and that 500 ml. of physiologic saline solution was added via the same infusion set at the second arrow. Blood or other types of fluids can be graphed in a similar fashion as given. Since two infusions in separate veins are not infrequently employed two lines are left for this purpose.

5. The next section represents analgesia, beneath which are four lines for graphic representation of plane or level of surgical anesthesia. The anesthetic level may be followed easily during the conduct of a given case if it is thus graphically recorded. These sections of the chart apply equally well to regional as to general anesthesia technics. Since definite stages and planes of general anesthesia are defined, the line on the graph goes lower as the anesthetic level becomes deeper. In regional or spinal anesthesia, the level of the body segment reached by the anesthetic at a given time can

similarly be recorded. The levels usually represented in spinal analgesia are: (1) below the twelfth thoracic segment; (2) below the seventh thoracic; (3) below the fourth thoracic or nipples; (4) above the fourth thoracic.

6. Immediately below this is the blood pressure, pulse, and respiration graph. To the left, is a box illustrating some of the more commonly used symbols. These are the symbols most commonly used by anesthesiologists to indicate various happenings graphically. It is a specialistic shorthand which is widely understood among members of the specialty. These are included to make interpretation easier to the uninitiated. It is not an all-inclusive list. When other symbols are used, they may be explained in the "remarks" column to the right of the graph.

The space allotted to "Remarks" is smaller than that ordinarily found on most record forms. This was made possible by the specific inclusion in the printed portion of a large part of the material which is frequently entered under "Remarks."

7. The bottom line of the graphic section indicates whether or not CO₂ absorption was employed. When "total" absorption is employed, i.e., with the closed system and soda-lime canister, the boxes in this line are completely cross hatched thus: . When a semi-closed system with partial absorption is used, the boxes are partially filled, thus: . In a completely open system, they are left empty.

8. Immediately beneath the graph is a line for indication, by means of a vertical arrow, the time of specific remarks, thus — " $\uparrow_1 \uparrow_2 \uparrow_3$ ". These are then entered in the "Remarks" column to the right of the graph, with the same key number being employed, e.g., "3—Neo-Synephrine 0.5 mg I.V."

9. The next two lines permit entry of induction, maintenance and emergence anesthetic agents with sufficient space for at least six different agents (important in these days of polypharmacy or balanced anesthesia). Below this, lines are provided for indication of anesthetic methods, means of airway maintenance, etc., and the usual identification of surgeons, anesthesiologists, postoperative diagnosis and operation performed.

The face of the card has 2 rows of punch holes along the right and left margins. The use of these will now be explained, beginning with the upper left-hand corner of the card. It should be emphasized at this point that all the information to be punched is contained in the central portion of the card. Thus, punching does not have to be done during the operative procedure, but can be done later either by the anesthetist himself or by a clerk or other person designated to do this work.

The holes along the left margin are punched as follows:

1. *Female*. The deeper hole is to be punched for *all* female patients. No punch is necessary for male patients.

2. *Spec. Int.* This section will be punched either deep or shallow, at the

direction of the particular department head, to indicate cases of unusual interest. It is an identifying means for quickly sorting the unusual from the more routine cases. These holes may be used for different purposes by different groups and at various times by the same group. For example, cases of unusual surgical interest such as rare tumors, extraordinary surgical procedures, etc., may be punched shallow, while cases of particular anesthesia interest may be punched deep.

3. *Anesthesia Time.* This section is used to indicate the time in hours that the patient was anesthetized. This, and the next section, operative time, are punched by numerals, as indicated—for less than $\frac{1}{2}$ hour, a shallow punch is made in the top space; for less than 1 hour, a deep punch is made. These sections provide a more extensive breakdown than some of the older records, in that $\frac{1}{2}$, 1, $1\frac{1}{2}$ hour cases can all be indicated. Cases longer than 5 hours are quite infrequent, and it is relatively unimportant to distinguish between a 6 and a 9 hour case because both are extremely prolonged surgical procedures and represent less than 3 per cent of all surgical cases.

4. *Operative Time.* Same as above, for anesthesia time.

5. *Special Studies.* This section is distinguished from section 2 above in that it is used for recording particular research studies rather than just interesting or unusual cases. Each study may be assigned a specific number which is punched as directed by the head of the department. As many as twenty different research studies can be carried on simultaneously and by means of reassignment of given numbers after a significant lapse of time the section can be used almost indefinitely. This section can also be useful in re-identifying cases to be studied after the card has been punched. For example, if it were desirable to study all cases of female patients below the age of 35 who had appendectomies performed with ether, this would ordinarily require sorting in four sections of the chart—sex, age, site of operation and anesthetic agent—whereas if such cases were arbitrarily assigned a number in the special studies section they could be found in a single sort. Other examples of special studies might be “cardiac arrest” cases, the “ether analgesia” technic, or hypnosis, which is again undergoing a period of increasing popularity. Each of these could be assigned a specific number in this section of the chart and the records could thus be made readily available. This section is coded as shown in Figure 8-2.

6. *Anesthetist.* The responsible anesthetist in each case will punch his own code number here. There is no necessity to learn more than this *one* code punch for each person. In the case of residents, the numbers can be reassigned when the resident leaves and the dates will distinguish one from another. This section is likewise coded as shown in (Figure 8-2).

7. *Preparatory Therapy.* This section is used for recording maintenance

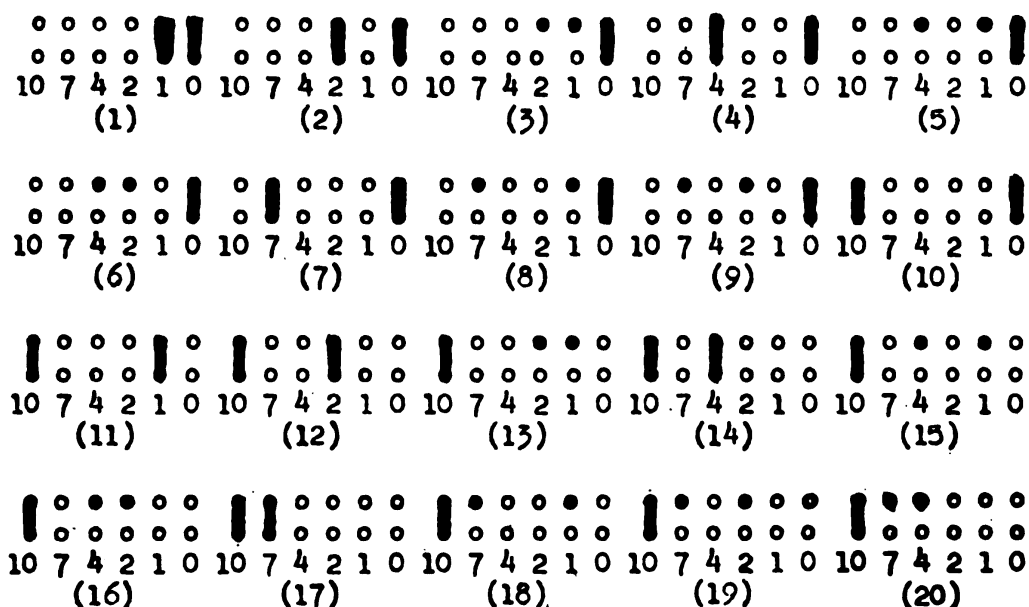


Figure 8.2. Explanation of coding Illinois E-Z Sort Systems anesthesia card. Special studies and anesthetist: The same code applies to both of these fields. These fields permit the direct extraction of any numeral from 1 to 20 inclusive, coded as shown above. Two sorting needles are required to extract numerals 1, 2, 4, 7, 10, 11, 12, 14, 17. Three needles are required to extract numerals 3, 5, 6, 8, 9, 13, 15, 16, 18, 19, 20.

therapy with drugs which may affect the conduct or outcome of anesthesia or surgery. This is distinguished from preanesthetic medication in that it represents *maintenance* therapy such as cortisone, digitalis, insulin, etc., which was employed preoperatively. This section is an innovation in the present card. To our knowledge it has not been employed previously and therefore its use may require somewhat more detailed explanation than other parts of the card. It was felt necessary to include such a section because the advent of a large number of newer drugs in recent years has complicated the administration of anesthesia. Many of these drugs such as the newer hormonal preparations can cause serious physiologic disturbances when anesthesia is subsequently administered, unless the anesthesiologist has been informed of their use and has taken proper precautions. It is not the purpose of this treatise to discuss individually the various therapeutic agents which may effect the conduct of anesthesia. Five of these have been listed on the card, and space has been left for seven others which may be added to the *armamentarium* at some future time. The five currently listed include digitalis, the narcotic drugs, antihypertensive and tranquilizing drugs, barbiturates, and endocrines. Other drugs such as radioactive isotopes, nitrogen mustards and similar antimalignancy agents, anabolic agents, and other types of chemotherapy which may alter the conduct of

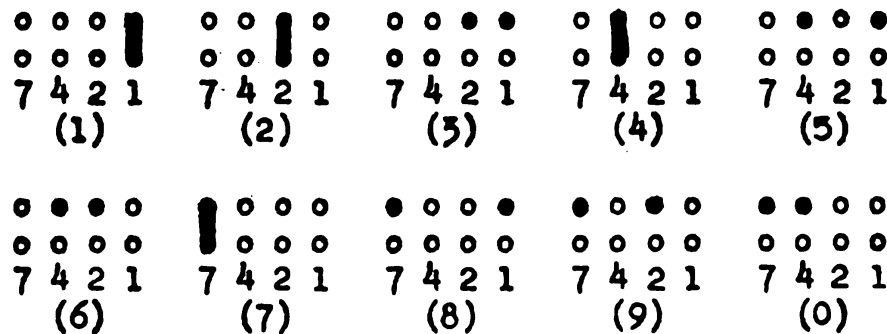


Figure 8-3. Explanation of coding. Year and physical status: The same code applies to both fields. The above field permits direct extraction of any numeral from 1 to 9 coded as shown above. A single sorting needle will extract numerals 1, 2, 4, 7. Two sorting needles are required to extract numerals 3, 5, 6, 8, 9, 0.

anesthesia may be assigned punches in the blank spaces provided in this section.

8. *Physical Status*. This is the standard A.S.A. physical status groups I thru VII. These are coded as shown in Figure 8-3.

Official American Society of Anesthesiologists' Classification

Non-emergency:

Class I. Patient has no systemic disease (though he may have a simple hernia or other nonsystemic disorder).

Class II. Patient has a mild systemic disease (such as moderate anemia, history of heart disease without symptoms or signs, mild diabetes under good control, etc.).

Class III. Patient has a moderately severe systemic disease which is not yet a threat to his life (such as heart disease with moderate symptoms, early carcinoma, severe diabetes under control, or such combinations as anemia plus mild diabetes, or chronic bronchitis plus dehydration, etc.).

Class IV. Patient has a severe systemic disease which is already a threat to his life (such as heart disease with failure, uncontrolled diabetes, or combinations of disorders that are a threat to his life, etc.).

Emergency:

Class V. Such emergency cases that would otherwise be classed in I or II.

Class VI. Such emergency cases that would otherwise be classed in III or IV.

Class VII. Moribund patients (such as the accident case that is in irreversible shock, severe bowel obstruction of many days duration, etc.) This category will rarely be used.

Mark the appropriate number in the *Physical Status Space*. In a good many anesthesia records the term "risk" or "operative risk" is used and a

designation is given for evaluation of this. In evaluating operative risk a good many factors must be considered, of which physical status is but one. In addition to the physical status of the patient, the kind and duration of the operative procedure must be taken into account. A patient who might be a normal risk for removal of an ingrown toenail under local anesthesia could conceivably be a very poor risk for major surgery such as a pneumonectomy.

The skill and experience of the surgeon should also be considered in evaluating the operative risk. The more skillful the surgeon is in performing the proposed surgical procedure, the less the operative risk. This also applies to the anesthesiologist. One who is versatile and well trained can carry a patient through much more difficult surgery with less risk than can the person of less experience.

Finally, duration of anesthesia and operation are important factors in risk. The surgeon who can perform a given procedure in one hour will encounter noticeably less morbidity than will one who requires four hours for the same operation. This then becomes a factor in risk.

It is because of the innumerable variables associated with "risk" and the changes that can occur in these from time to time even during the conduct of an anesthetic that the authors decided it was of little value from a statistical viewpoint. We, therefore, chose to use "physical status" which is always the same for a given patient on a given day regardless of whether or not operation is contemplated.

9. *Age of Patient.* This section differs only slightly, but significantly, from previous cards. The ages recorded are in years. The first punch (—1) covers all infants from newborn to 1 year of age. The second (—5) covers all patients from 1 to 5 years. These pediatric cases have largely been neglected in previous recording by decades, but from the point of view of the anesthesiologist the young child presents very different problems from the older child or adult. It was therefore decided to include this group in a separate punch. From here on age is classified by decades up to age 70. Patients over 70 years old are generally classified as geriatric cases so that a further breakdown was not considered important.

10. *Year.* This is a coded section, but again, only one coding is used by all individuals concerned, for an entire year, so that cards can be pre-punched prior to use if so desired. This is coded in the same manner as "physical status" above.

The right-hand side of the card requires somewhat less explanation for its proper utilization:

1. The eight upper portions of this side of the card refer to *site of operation*. These should be self-explanatory. They are similar to the designations employed in the Chicago Keysort and similar edge-punched cards.

2. Position, again self-explanatory—refers to the position of the patient during the operative procedure. Nine of the most common positions are indicated with specific punches and a blank is provided for unusual positions not ordinarily employed.

3. *Level or Plane*. A shallow punch refers to plane of surgical general anesthesia and deep punch refers to height or level of spinal or regional anesthesia.

4. The next sections refer to type of operative procedure—diagnostic, therapeutic, definitive, palliative, major or minor. Conceivably a given operation might be—therapeutic, definitive and major. These have all been available for indication on previous record forms but the authors feel that the arrangement has been such that they were not utilized to their fullest capabilities. Here, they have been placed in a single section of the card and can easily be punched.

5. *Anesthesia or Analgesia*. This refers to the type of services rendered by the anesthesiologist. It will usually be the former, but in such cases as burn dressings, painful examinations, first and second stages of labor, etc., analgesic technics may be employed. Analgesic technics may also be used in therapeutic nerve blocks for pain, etc.

6. *Death*. This section classifies deaths as preventable or nonpreventable, and as to whether or not anesthesia contributed. Accurate use of this section requires a full discussion of the case by all concerned prior to punching. It should not be abused by punching without mature consideration of all deaths. If an autopsy has been performed its results should be known to the anesthesiologist before this section is punched. If an autopsy has not been performed the entire clinical record, including the anesthesia chart, should be carefully reviewed and the department should sit in judgment as to whether death was preventable or non-preventable and as to whether or not anesthesia contributed. It is rare that more than a few deaths would occur in a month's time in any single department. It is logical therefore to suggest that all deaths be discussed in a monthly department meeting and that the cards be punched after the meeting.

7. Sections "B" and "C" are unassigned and may be utilized in any manner desired by a specific department head. For example "cardiac arrest" cases and such things as neurological complications of spinal anesthesia, specific drug reactions, cases of surgery performed by distinguished surgeons visiting the hospital or any other desired data might be recorded here.

The reverse side of the card (Figure 8-4) is placed in a vertical position so that it can also be punched along the right- and left-hand margins. Four rows of punch holes are provided along each margin. The upper central portion of the card provides a number of blank lines for entering preopera-

tive and postoperative summaries. Starting in the upper left corner the punches on this side of the card are as follows:

Left Hand Margin (Punched)

Pre-Anesthetic Medication
Relaxants & Their Antagonists
Vasopressors
I.V. Therapy
Pre- and Post-operative Complications

Right-Hand-Margin (Punched)

General Anesthesia Techniques
Endotracheal Techniques
Spinal Techniques
Regional Techniques
Hibernation, Hypothermia, etc.
Regional agents
General Anesthetic Agents
Anesthesia Complications

1. *Pre-anesthetic Medication*. Spaces are provided for drugs, route of administration, and effect. The agents and techniques usually employed are printed on the card, and 13 blanks have been provided for additional ones.

2. *Relaxants*. Here again the usual ones are listed with 7 blanks provided for additional agents.

3. *Vasopressors*. This lists the usual drugs plus blank space, with space for recording dosage, technique of administration, and reason for its use.

4. *I.V. Therapy*. This section permits recording of blood, plasma volume expanders, electrolytes, sugars, etc., with 14 blank spaces plus the more commonly used agents.

5. The next section extends to the bottom of the left-hand side of the card. It permits punching of all pre- and post-anesthetic complications plus recording of specific complications. An ingenious punching device permits punching either preoperative or postoperative complications, or both, plus an additional line for deaths. The first hole represents preoperative complications; the second hole represents postoperative complications. If both pre- and postoperative are present in the same category, the third hole is punched, and if death ensues, the fourth is punched. For example, a patient with pulmonary TBC, active, would have the first hole punched under *Resp. Major*. If this same patient then developed atelectasis postoperatively the third hole would be punched. Whereas had the patient developed atelectasis without the pre-existing TBC the second hole would have been punched.

Preoperative complications may be punched at the time of the preoperative visit to the patient. If the same patient later develops postoperative complications this could then be easily recorded by punching the third

Figure 8-4. Illinois E-Z Sort anesthesia punched card (back).

instead of the second hole. If death later ensued it could still be indicated by punching the fourth hole. The specific complications are not individually punched but the more common ones are grouped according to systems and printed on the lines opposite the system punched. For example, under the heading "NEUR DISEASES" is listed "psychosis." If a patient developed a psychotic episode postoperatively this section would be punched in the

second hole and psychosis circled with a pencil with the notation "P.O." after it. In the event the patient was a known epileptic and the anesthetic resulted in no aggravation of his epilepsy, the first hole would be punched and the word EPILEP circled with the notation "Pr.OP" after it.

Each of the sections in this grouping is handled in a similar manner. The specific complications are circled or written in and the applicable punch is made for the group under which a specific complication is classified.

It is true that if one were searching for cases of a specific complicating disease, for example—"Nephritis"—it would be necessary to extract all cards for "GU diseases" and then hand sort for "Nephritis." The number of cards to be hand sorted would not ordinarily be unwieldly and in the event that it were known in advance that a certain specific disease or condition was to be studied, the punches "B" or "C" on the face of the card could then be utilized for this purpose.

For example, in a mental hospital it is to be expected that a good proportion of all patients would have neuropsychiatric disorders. In the event it was desired to study all epileptics in such a hospital, "B" punched shallow on the face of the card (lower right corner) could be assigned to epilepsy and this would tend to simplify sorting. Punch "C," if desired, might likewise be assigned to manic depressive psychosis.

The right-hand border on the reverse of the card represents technics and agents of anesthesia plus complications occurring during administration.

1. The topmost section includes technics of general anesthesia. The four holes at each technic represent, respectively, (1) induction, (2) maintenance, (3) emergence, and (4) supplemental. (The latter refers to technics used to supplement the maintenance with another agent (for example, intravenous supplementing inhalation).

2. The second section classifies the types of tracheal intubation.

3. The next two sections include technics of spinal and regional blocks.

4. Anesthetic agents are listed next. The regional agents are given first, with suitable blank spaces for the newer drugs. The agents for general anesthetics are given, provided with classifying punches, viz: (1) primary, (2) induction, (3) supplementary, (4) emergence.

5. Finally, a section is provided in the lower-right hand corner for recording anesthetic complications and the time of their occurrence during induction, maintenance, emergence or following premedication.

The reverse of the card can be marked and punched at any time after the case has been concluded. We believe this part of the card will furnish the most valuable statistical information.

We also feel that much needless searching of records for statistical data can be readily avoided by the use of this card. For hospitals interested in less detailed information, the face of the card presents a readily available medico-legally complete record and this side alone may be used.

References

1. Tovell, R. M., and Dunn, H. L., "Anesthesia Study Records," *Anesthesia & Analgesia* 11, 37-41 (Jan.-Feb. 1932).
2. Rovenstine, E. A., "Method of Combining Anesthetic and Surgical Records for Statistical Purposes," *Anesthesia & Analgesia*, 13, (May-June 1934).
3. Nosworthy, Michael, "A Method of Keeping Anesthetic Records and Assessing Results," *Brit. J. Anesth.*, 17, 160-179 (July 1943).
4. Pender, John W., "A Combined Anesthesia Record and Statistical Card," *Anesthesiology*, 7, 606-610 (Nov. 1946).
5. Conroy, W. Allen, Cassels, W. H., and Stodsky, Bernard, "The Chicago Keysort Anesthesia Record," *Anesthesiology*, 9, 121-133 (March 1948).
6. Sadove, M. S., and Levin, M. J., "The Illinois E-Z Sort Anesthesia Record Card," *Anesthesiology*, 19, 178-187 (March-April 1958).

Chapter 9

PUNCHED CARDS AS AIDS TO QUALITATIVE CHEMICAL ANALYSIS BY SPECTRAL METHODS*

L. E. KUENTZEL

Wyandotte Chemicals Corporation

Wyandotte, Michigan

Introduction

Qualitative chemical analysis employing physical methods, stripped of confusing but necessary trimmings, almost always involves the measuring or more or less distinctive physical properties of a compound and comparing them with similar physical data obtained from compounds of known purity by the same reproducible methods. Many such physical properties are simple to measure and have been used for identification purposes for some time. Among these are such properties as the boiling or melting point, index of refraction, density, crystalline form and optical activity. Moreover, it is easy to record and tabulate such data for the preparation of tables of standard data for identification purposes. However, the physical properties of compounds as measured by the spectral methods of absorption spectroscopy, x-ray diffraction, mass spectroscopy, raman spectroscopy and others are of such complex nature that they cannot be reduced to a simple number. In such cases one must compare complex sets of data with standard sets of similar data to identify the unknown material being analyzed. Although the very complexity of such sets of data provides highly desirable details for correlation purposes it does present problems in connection with tabulating, sorting, comparing and distributing the data in the normal course of qualitative analysis. This has led to an increasing use of notched-type punched cards. Also, more recently, under pressure of rapidly accumulating tens of thousands of such sets of standard data the use of International Business Machine cards and sorting equipment has become extremely helpful. This chapter describes the practical application of punched cards, both the

* Thanks are due the American Society for Testing Materials for permission to reproduce in this chapter its copyrighted Codes and charts pertaining to ASTM-Wyandotte punched cards for indexing spectral absorption data. Thanks are due the Consolidated Electrodynamics Corp., Pasadena, California, for permission to reproduce in this chapter its copyrighted mass spectrum card.

notched or Keysort type and the punched or IBM type, to the problems of indexing and correlating complex sets of physical data for qualitative analytical purposes and presents the codes and instructions for the most widely used systems.

General

There is no need in this chapter to discuss at length the fundamentals of preparing and using punched cards. Chapters 2 and 3 provide descriptions of such operations. Therefore, attention will be paid to particular examples of the uses of such cards. Before discussing such applications, a brief comparison of the relative advantages and disadvantages of the two types of cards in this particular field will assist the reader in evaluating the most desirable method for his specific problems.

Hand-Sorted Cards. The chief advantage of the hand-sorted card lies in the fact that standard data for which one is searching may be printed or written on it so that once the proper card is located much of the pertinent information may be read directly from it. Also, sorting operations can be carried out by hand with uncomplicated and inexpensive equipment. The notching of such cards requires inexpensive equipment and errors are easily corrected. However, sorting such cards requires much physical manipulation and the handling of 5,000 or more cards with needles becomes a real chore. The number of notch or code positions is limited by the size of the card and the larger the card the more difficult it is to handle. Finally, such cards are expensive to prepare and reproduce and are subject to rather rapid deterioration with much use.

Machine-Sorted Cards. The chief advantages of this type of card are the greatly increased coding possibilities on small cards and the ease and accuracy with which they may be handled. On a $3\frac{1}{4}$ by $7\frac{1}{4}$ inch IBM card there is space for 960 direct code punches. It would take a notched card over four feet square to provide the same information. Also, machine sorting is so effortless that tens of thousands of cards pose no special problems. IBM cards are inexpensive and easy to prepare and reproduce. On the other hand, only a limited amount of information can be printed on such cards. It thus becomes necessary to use the cards merely to index the standard data rather than to provide a means of recording and filing such data. This necessitates the maintenance of two files. The file of detailed standard data may be kept in whatever form it is obtained, i.e., film strips, recording charts, published curves or tabulations; the actual searching for wanted portions of this data is done with punched cards which bear serial number references to the location of the standard data. IBM cards necessitate the use of mechanical equipment. While this can be quite expensive in some cases, most of the applications described in this chapter require only a sorter.

It has been widely agreed that machine sorting is best for universal searches where many thousands of cards are involved and the data are complex, and that hand sorting is most efficient for small numbers of cards indexing limited ranges of data. Many laboratories find it convenient to use both types. In this discussion that follows, examples will be presented showing how both hand-sorted and machine-sorted cards are applied in handling data in several fields of physical analytical chemistry.

Infrared Absorption Spectroscopy

The physical data obtained from a compound by an infrared absorption spectrograph are complex and are usually represented by a plot of the per cent transmittance versus the wavelength or frequency of the infrared radiation. Modern spectrographs produce such a plot or spectrogram automatically but the shape and size of the spectrogram varies with the make of the instrument. Much data is still hand plotted for publication. Here, then, is a case where standard data obtained from compounds of known purity, with which the unknown spectrogram must be compared for identification purposes, exist in the form of thousands of spectrograms, plotted in several coordinate systems, in a variety of shapes and sizes, and located in many books,¹⁻³ journals and catalogs.⁴⁻⁶ ^{6A} This makes the physical matching of standard and unknown data for qualitative analysis unduly time-consuming. Consequently, a great deal of effort has gone into the problem of adapting punched card techniques to the solution of the correlating problem.

The fact that much of the data in an infrared spectrogram can also be correlated with specific structural features of the compound from which it was obtained has made it desirable to include in punched card systems the details of chemical structure or groups which are singly responsible for specific segments of the spectrogram. In order to anticipate future correlations of structure and absorption data, the chemical structure codes usually have been as embracing as space limitations on the various cards permit. In some systems it has been desirable to include important elements, melt-

¹ Barnes, R. B., R. C. Gore, U. Liddle, and V. Z. Williams, "Infrared Spectroscopy," New York, Reinhold Publishing Corp., 1944.

² Dobriner, K., E. R. Katzenellenbogen, and R. N. Jones, "Infrared Spectra of Steroids," New York, Interscience Publishers, 1953.

³ Randall, H. M., R. G. Fowler, N. Fuson, and J. R. Dangel, "Infrared Determination of Organic Structures," New York, D. Van Nostrand Company, 1949.

⁴ "Catalogue of Infrared Spectral Data," American Petroleum Institute, Research Project 44, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

⁵ "Catalogue of Infrared Spectrograms," Samuel P. Sadtler and Son, Inc., 1517 Vine Street, Philadelphia 2, Pennsylvania.

⁶ National Research Council Committee on Spectral Absorption Data, Mr. J. J. Comeford, Secretary, National Bureau of Standards, Washington 25, D. C.

^{6A} "The DMS System," London, Butterworths Scientific Publications, 1956.

ing or boiling points, and other physical characteristics of the compounds in the information coded into the cards. This provides greater flexibility in card sorting when such data are available.

Hand-Sorted Cards. A great many laboratories have developed codes for and make use of such cards to meet their particular needs. It will not be possible to describe all of these. However, many of them are quite similar since the objective in all cases is much the same. The system currently distributed by the National Research Council Committee on Spectral Absorption Data in collaboration with the National Bureau of Standards represents the best features of some eighteen such systems. It was developed by a Punched Card Committee appointed by the 1948 Symposium on Molecular Structure and Spectroscopy at Ohio State University. Because of the wide acceptance of this card and system, the description which follows is sufficiently complete to provide working instructions for the use of the cards.

The edges of the N.R.C.-N.B.S. card (Figure 9-1) are divided into four fields which provide for coding and sorting the following data:

- (1) Positions of Major Absorption Bands
- (2) Melting or Boiling Point
- (3) Molecular Functional Groups
- (4) Number of Carbon Atoms

Fields (1) and (2) permit entry to the file in the case of unknown compounds while (3) and (4) permit entry in the case of known compounds.

(1) *Wavelength-Wave Number.* This field includes from 2.70 (3700 reciprocal centimeters) to 40.0 microns (250 reciprocal centimeters) in nonlinear intervals. The spectral range covered by each notch position is printed on the card and the coding of each major band is achieved by notching the card between the printed values which encompass the major band. Only the relatively strong absorption bands of a given spectrogram are coded into the card, and sorting these corresponding fields serves to isolate those cards indexing compounds whose spectra match the one sought for. A comparison of the unknown spectrogram with those appearing on the back of the card (Figure 9-2) serves to establish the identification.

(2) *Melting or Boiling Point.* A melting or boiling point is notched into the field on the right of the card. The familiar 1, 2, 4, 7, SF system for notching numbers is used and space is provided for three-digit, whole-number values. A small field adjacent serves to indicate whether the notched value is a melting or a boiling point as well as to indicate negative values when necessary.

(3) *Molecular Functional Groups.* Holes numbered 1 through 31 are a direct sorting code for molecular functional groups, which were chosen because they are known or suspected to have characteristic absorption in the

NAME		SERIAL NO.	STRUCTURAL FORMULA					
N-BUTYL ETHER		33	$\text{CH}_3-(\text{CH}_2)_3-\text{O}-(\text{CH}_2)_3-\text{CH}_3$					
(DI-N-BUTYL ETHER)		$\text{C}_8\text{H}_{18}\text{O}$						
RENS. P. I. LIQ. 2-15 μ								
B.P. 142°C								
REFERENCE NUMBER	STATE	SAMPLE THICKNESS	PRESS. CONC. SOLVENT	RANGE	DATE	PRISM GRATING	TEMP.	REMARKS
A.P.I. 1234,								
3315-1391					1939			EFFECT AS SOLVENT ON O-D BAND
3315-1392					1939			EFFECT AS SOLVENT ON N-CL BAND
3315-1394					1939			EFFECT AS SOLVENT ON O-H BAND
9434-0501	L	.025mm		720-750	1950	NACL		BAND WL.
0461-0441					1943			ANAL. FOR IN BUTYL ACRYLATE
0461-0441				800-1500	1943			SPECTRUM
2515-0242	L	11mm		TO 2.5 μ	1924	GLASS		SPECTRUM
2515-0291	L	1mm 40mm		0.7-2.5 μ	1929	QTZ.		C=O STUDY
1979-0251	L	1.5mm		2-7 μ	1925	NACL		SPECTRUM

NRC-NBS INFRARED COMPOUND CARD

Figure 9-1. Front of NRC-NBS Compound Card.

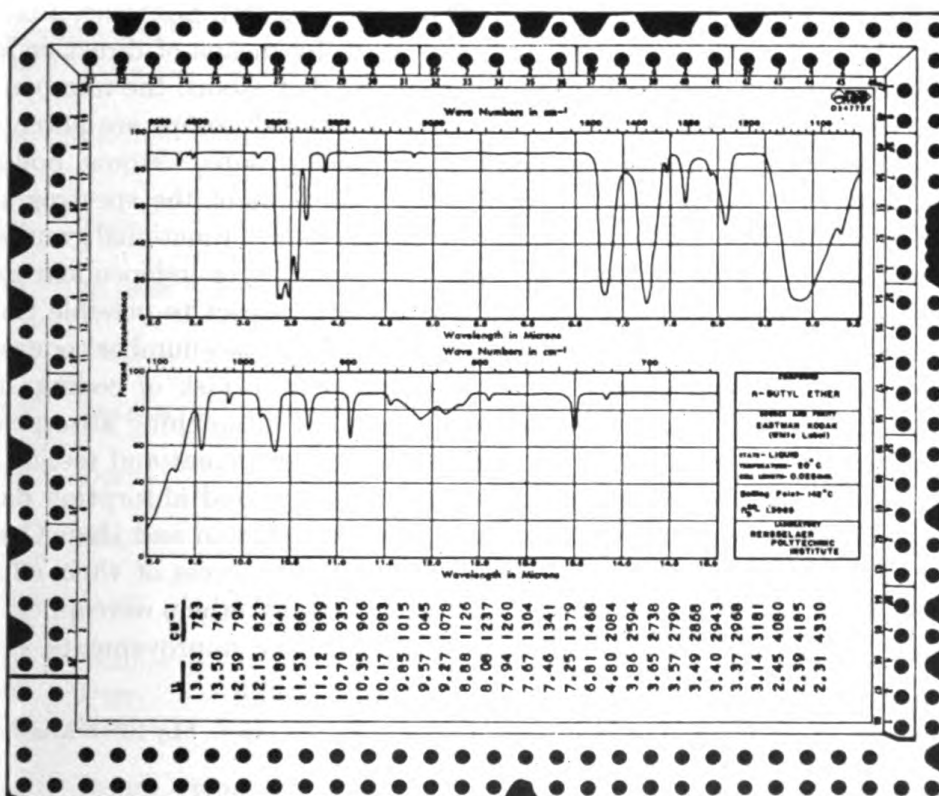


Figure 9-2. Back of NRC-NBS Compound Card.

infrared region. Both deep and shallow notching is used in this field and abbreviations of the corresponding molecular group are printed under each hole on the card. Table 9-1 gives a complete description of each abbreviated term.

(4) *Number of Carbon Atoms*. This field provides for recording such information about the compound with the 1, 2, 4, 7, SF code and sufficient space is available to indicate 39 or less atoms. This information, together with sorts on functional groups, serves to help isolate the card indexing a given, known compound and to eliminate unwanted cards when partial information is available about an "unknown" compound.

Space is provided on the face of the card for the name of the compound, structural and empirical formulas, range of the spectrogram (printed on the back of the card), state of the sample, purity, serial number, etc. A tabular list of references to other published spectra of and information about the particular compound is also provided. The back of the card carries the infrared absorption spectrogram of the compound and a list of the spectral positions of the absorption bands. These cards are available from the National Bureau of Standards.⁶

Mention should be made of a few other notched card systems which have worked quite well. One such system has been developed by the Shell Development Company.⁷ In order to provide more notch positions, a large card, 8½ by 11 inches, is used. Into this card, by means of direct codes, the spectral region covered, the number and kind of atoms, the number of carbon atoms and the number and kind of functional groups are notched. Two-number codes are used to notch "functional groups," atoms, boiling point and wavelength positions of major absorptions of the spectrogram into the rest of the card. Unique is the use of very small "functional groups," the indication of whether these groups are interacting or independent and whether they are acyclic, cyclic or exocyclic with respect to possible rings in the compound being coded. Although the use of the two-number code and the small "functional groups" adds somewhat to the task of coding and sorting the data, it does permit greater resolution in notching absorption positions and structural detail. Pioneers in the development and testing of punched card methods for indexing and sorting infrared absorption data must include workers at Dow Chemical Company,⁸ Rohm and Haas Company⁹ and United States Rubber Company.¹⁰ The success of their effort is reflected in the N.R.C.-N.B.S. card of today. Many others were quick to support these efforts and to supply modifications and improvements. Par-

⁷ Brattain, R. R., personal communication, February, 1952.

⁸ Wright, N., "Infrared Spectroscopy," *Am. Chem. Soc. Abst.*, **6L**, 108th Meeting, New York City, (September 1944).

⁹ Stroupe, James D., personal communication, December, 1952.

¹⁰ Hampton, R. R., personal communication, December, 1952.

TABLE 9-1. LIST OF ABBREVIATIONS FOR NATIONAL
RESEARCH COUNCIL INFRARED PUNCH CARDS

Compound card—functional group code			
No.	Position	Code	Class or Sub-class
1	Shallow	Cl	All cpds. containing Chlorine.
1	Deep	Cl _n	Cpds. with more than one Cl atom per molecule.
2	Shallow	HAL	All cpds. containing F, or I.
2	Deep	Br	All cpds. containing Br.
3	Shallow	CONJ	All cpds. with conjugated nonaromatic double bonds.
4	Shallow	C≡C	All cpds. with an acetylenic bond.
5	Shallow	HC	All hydrocarbons, incl. C isotopes, but not H isotopes.
5	Deep	SAT	All saturated hydrocarbons, incl. C isotopes, but not H isotopes.
6	Shallow	POL	All polymers.
6	Deep	CO-POL	All copolymers.
7	Shallow	C=C	Cpds. with aliphatic C=C double bonds, not incl. in 7 deep and 8.
7	Deep	R ₁ C=CHR	Cpds. containing the aliphatic double bonds of types R _a R _b C=CHR _c and R ₂ C=CR ₂ , R same or different but not H.
8	Shallow	=CH ₂	Cpds. with aliphatic double bonds of the type C=CH ₂ .
8	Deep	R ₁ C=	Cpds. with aliphatic double bonds of the type R _a R _b C=CH ₂ .
9	Shallow	AR	All cpds. containing an aromatic carbon ring not incl. in 9 deep, 10, 11, 12. Fusing with another ring constitutes substitution.
9	Deep	MONO	All cpds. containing a mono-substituted aromatic carbon ring.
10	Shallow	PARA	All cpds. containing para di-substituted aromatic carbon rings.
10	Deep	ORTHO	All cpds. containing ortho di-substituted aromatic carbon rings.
11	Shallow	META	All cpds. containing meta di-substituted aromatic carbon rings.
11	Deep	UNSYM	All cpds. containing unsymmetrical tri-substituted aromatic carbon rings.
12	Shallow	SYM	All cps. containing symmetrically tri-substituted aromatic carbon rings.
12	Deep	VIC	All cpds. containing vicinally tri-substituted aromatic carbon rings.
13	Shallow	HET	All cpds. containing heterocyclic rings.
13	Deep	HET-N	Heterocyclic cpds. with nitrogen in the ring.
14	Shallow	ALCY	Cpds. with carbon rings other than aromatic rings.
14	Deep	ALCY-Δ	Cpds. with C=C in a carbon ring other than an aromatic ring.

Table 9-1. *Continued*

Compound card—functional group code			
No.	Position	Code	Class or Sub-class
15	Shallow	OH	All cpds. containing OH except C=C(OH); including hydrates.
15	Deep	ALC	All alcohols (excluding phenols).
16	Shallow	AC	All carboxylic acid anhydrides, and salts.
16	Deep	ACID	All carboxylic acids.
17	Shallow	AL	All aldehydes.
17	Deep	KE	All ketones.
18	Shallow	EST	All esters of carboxylic acids.
18	Deep	ACET	Esters of acetic acid.
19	Shallow	ETH	All oxygen C—O—C ethers.
19	Deep	METH	Methyl oxygen ethers.
20	Shallow	CO	Cpds. with C-to-O bonds not included in No. 15 Nos. 19 and 20 deep.
20	Deep	AMD	Acid amides and N-substituted acid amides. —C=O(NH ₂), —C=O(NHR), —C=O(NR ₂).
21	Shallow	CS	All cpds. with C—S bonds.
21	Deep	C=S	All cpds. with C=S bonds.
22	Shallow	S	All other cpds. containing sulphur not in 21 and 22 deep.
22	Deep	SO	All cpds. containing S-to-O bonds including inorganic radicals.
23	Shallow	CN	All cpds. with N-to-C bonds, including tertiary amines but not 20 deep, 23 deep and 24 deep.
23	Deep	C≡N	Nitriles.
24	Shallow	NH	Secondary amines, not acid amides.
24	Deep	AMIN	Primary amines.
25	Shallow	NO	All cpds. with N-to-O bonds, except 25 deep.
25	Deep	NO ₂	All cpds. having $\text{N} \begin{array}{c} \text{O} \\ // \\ // \\ \text{O} \end{array}$ groups.
26	Shallow	N	All cpds. containing nitrogen not covered elsewhere.
26	Deep	N—N	Cpds. containing N-to-N bonds.
27	Shallow	SI	All cpds. containing silicon.
27	Deep	—	Not assigned.
28	Shallow	NON-M	All cpds. which contain non-metals not covered elsewhere.
28	Deep	P	Cpds. containing phosphorus.
29	Shallow	MISC	Miscellaneous cpds. not included elsewhere.
29	Deep	MET	Organo-metallic cpds.
30	Shallow	ISOT	Cpds. containing isotopes other than D and T
30	Deep	DEUT	Cpds. containing deuterium and tritium.
31	Shallow	—	Unassigned.
31	Deep	—	Unassigned.

ticular mention should be made of a unique approach to the problem made by a worker at Cornell Medical College,¹¹ where the system devised for cataloging x-ray diffraction data based on the three strongest lines¹² was applied to infrared absorption data. A new and comprehensive system for coding chemical structure and spectral absorption data into edge-notched cards has been proposed by Thompson^{12A} which forms the basis of an extensive catalogue of data being issued from England.^{6A} These data are being coded into IBM cards also.

Machine-Sorted (IBM) Cards. An early application of IBM cards and equipment to the problems of sorting and correlating infrared absorption data for purposes of qualitative analysis was developed by Wyandotte Chemicals Corporation.¹³ The American Society for Testing Materials* has since assumed the responsibilities connected with the development and maintenance of this and other indexing systems originating at Wyandotte, and which are being used by an increasing number of laboratories. This provides for the perpetuation and orderly modification of the systems in the interests of a greater number of people, provides a mechanism for supplying the cards and codes through a single agency, and furnishes the manpower to insure greater coverage and accuracy of the data issued. A research fellowship at the National Bureau of Standards has been established by the American Society for Testing Materials to provide for the preparation of master cards coded by members of the Standard Data Subcommittee of A.S.T.M. Committee E-13. A description of the Wyandotte-A.S.T.M. system for indexing infrared absorption and chemical structure data, together with pertinent codes and charts, follows; other system will be discussed later in this chapter.

Wyandotte-A.S.T.M. infrared punched cards are designed to facilitate the sorting of spectral absorption data for the purpose of matching spectrograms in qualitative analysis and to aid in correlating chemical structure and absorption band positions. The cards are merely an indexing system which enables one to make rapid, accurate searches by machine and inexpensively to obtain and maintain ready reference to all published spectra.

¹¹ Clark, C., "Cataloguing of Infrared Spectra," *Science*, 111, 632-633, (June 9, 1950).

¹² Hanawalt, J. D., H. W. Rinn, and L. K. Frevel, "Chemical Analysis by X-Ray Diffraction," *Ind. Eng. Chem., Anal. Ed.*, 10, 457 (1938).

^{12A} Thompson, H. W., "The Documentation of Molecular Spectra," *Journal of the Chemical Society*, pages 4501-4509, (1955).

¹³ Kuentzel, L. E., "New Codes for Hollerith Type Punched Cards," *Analytical Chemistry*, 23, 1413-18 (1951).

* Thanks are due the American Society for Testing Materials for permission to reproduce in this chapter its copyrighted Codes and charts pertaining to ASTM-Wyandotte punched cards for indexing spectral absorption data.

Through the cooperation of the Technical Information Division of the Battelle Memorial Institute, the National Bureau of Standards and members of A.S.T.M. Committee E-13, all publications are monitored for infrared spectra which, together with those issued by regular publishers of spectra, are coded and incorporated into the card files. Cards, indexes and instruction booklets¹⁴ may be obtained from the A.S.T.M.

For reason of simplicity and economy of sorting time, the system makes widest possible use of direct codes and requires the use only of a standard Type 82, 80-1 or 80-2 Sorter available from International Business Machines, Incorporated. Although considerable information about the operation and use of common pieces of IBM equipment is available elsewhere in this book, the manipulation of the sorter as it applies to the handling of these cards will be reviewed briefly.

The IBM card bears eight vertical columns of numbers from 0 through 9 which mark the positions where small rectangular holes will be punched to record data. Above the 0 position in each column are two unmarked over-punch positions referred to as "x" and "y" positions, with the latter being uppermost. Numbers can be punched in directly. Thus, the number 457 can be recorded by punching the digits 4, 5 and 7 in any three adjacent columns. Letters of the alphabet require a special two-punch-per-column code. When used for these purposes, only one digit or letter can be punched into a given column. However, a direct code can be set up to relate any given punch position to any particular item. Thus, a punch at number 7 in column 42 can always mean that there is an —OH functional group in the compound described by the card. With this type of coding it is possible to record from one to 12 items independently and at the same time in a single column of a card. Moreover, by making use of the selector switches, any item so coded may be searched for singly regardless of the number of other items coded into the same column. In sorting operations, the sorter senses one column at a time. As the cards pass over a metal cylinder, a small metal brush makes electrical contact with the cylinder through the holes punched in the cards. This causes the card to be shunted into a pocket corresponding to the number at which the punch is made. If there is more than one punch in the column, the machine sends the card to the pocket bearing the highest number so punched. Positions "x" and "y" rate below 0, with the "y" position being the last sensed by the machine. However, if any one or more of the 12 selector switches are moved to the "off" position, the machine will ignore punches at those positions and respond to any remaining punches, and if there are none, it will then send the card to the

¹⁴ "Codes and Instructions for Wyandotte-A.S.T.M. Punched Cards Indexing Spectral Absorption Data," the American Society for Testing Materials, 1916 Race Street, Philadelphia 3, Pennsylvania, 1954.

Figure 9-3. Wyandotte-A.S.T.M. Infrared Data Card.

The Wyandotte-A.S.T.M. card indexing infrared absorption data (see Figure 9-3) is divided into the following areas for coding purposes:

- All the data available from one compound and covered by the codes are punched into one card. In order that this description of the card and system may serve to instruct one in the use of the cards, codes and examples are included.

(1) *Infrared Absorptions.* The coding of absorption band positions is done in terms of wavelength in microns. As a general rule, all bands having an absorbance ratio with the strongest band in the spectrogram of 1:10 or more are coded. From columns 1 through 25, the column number is taken as the whole number value of the absorption band and the fractional part is rounded off to the nearest tenth of a micron and coded by punching the digit into the appropriate column. Thus, a band at 7.38 microns would be coded by a single punch at the 4 position in column 7. From 25 to 50 microns the punching resolution is 1.0 micron and coding is achieved by punching the units values into columns headed by appropriate tens values. Thus a value of 34.8 microns is coded by a single punch at the 5 position in column

27 which is headed by the number 30. Punches at the "x" position in this section of the card serve to indicate that there are no data available for the region covered by the particular column or columns involved. Finally, a "y" overpunch is used to code the position of each very strong band which may be expected to persist in the spectrum of a considerably diluted sample of the material.

Sorting operations for coded absorption data follow one of two possible courses depending whether the unknown material is relatively pure or is a mixture of several compounds in roughly equal concentrations. In the former case, positive sorting on the strong band positions is appropriate while the latter case involves initial negative sorting before making positive sorts. Each method is described briefly.

Positive sorting on coded absorption band positions seeks to segregate cards bearing code punches corresponding to those of the major absorptions in the spectrogram of the unknown material. The band code positions are sorted for one at a time, beginning with the most characteristic or unique band. This tends to eliminate the greatest number of cards in the initial sort and is usually achieved by sorting on the longest wavelength band first. By making use of the selector switches, the sorter can be set to search for cards having either specific individual punches or a range of consecutive punch positions. The latter operation permits sorting for a narrow range of absorption band positions simultaneously when the exact location of the band peak is not apparent in the spectrogram and thus avoids missing the wanted card. Each sort is made upon the small residue of the previous sort and the number of cards diminishes rapidly so that a single card frequently results in from three to five sorts even when starting with several thousand. Each card bears a serial number which directs the searcher to the location of the standard infrared spectrogram from which identification of the unknown spectrogram may be made.

In spectra of mixtures of compounds, it is not known which bands belong to chemical individuals so that sorting of the type just described is not feasible. One seeks first to eliminate all cards which have absorption bands in wavelength regions where the spectrogram of the unknown does not have any, since none of these materials could possibly be a component of the mixture. This is accomplished by negative sorting of the "transparent" regions of the unknown spectrogram, i.e., setting the sorter to segregate into pockets all cards that *do* have bands in these regions, then discarding them and keeping the cards which accumulate in the reject pocket. Use of the "y" overpunches facilitates negative sorting since the relatively weak bands of a minor component do not interfere with such sorts in otherwise transparent regions. The relatively small deck of cards that results from the negative sorting operation is then subjected to positive sorting on the

band positions, following a systematic trial and error schedule that considers the possible combinations of bands that may characterize an individual component of the mixture. Here, again, it is expedient to begin with the long wavelength bands and test-sort all possible combinations until an identification is made. When this is done the bands belonging to that compound are eliminated from consideration and the process continued on the remaining bands. Although it sounds complicated, the actual time involved is small and the process is easy to carry out. A typical example of the analysis of a three-component mixture starting with over 3,000 cards required about 18 minutes of negative sorting and 3 minutes of positive sorting on the residue deck to identify the three unknowns.

(2) *Chemical Classifications.* The philosophy behind the development and use of the Chemical Classification Code attempts to divorce the complexities of the names and chemistry of organic compounds from the codes and coding operations used to characterize them. It is not intended that each such characterization be unique for each different molecule since the purpose of the code is to provide a means of segregating compounds into related groups. Coding is based entirely upon a detailed structural formula and a recognition of the "code units" which make up the formula. These code units, in many cases, are the same as familiar reactive groups or radicals that enter into the chemistry and naming of organic compounds, but such names and chemistry as may be associated with the code units must not restrict the use of the unit wherever applicable under the rules presented.

Table 9-2 relates the code punch positions on the card in terms of column and row numbers to the coded items of structural features used to characterize compounds. References to these punch positions are made by giving the column number, then a dash followed by the punch positions. For example, 32-0, 2, 4 indicates punch positions 0, 2 and 4 in column 32 which code the presence of elements oxygen, sulfur and chlorine, respectively. Table 9-3 lists some examples of the types of compounds which the code units index. Following is a column-by-column discussion of the chart and a presentation of the rules set up to insure uniformity in coding.

Column 32 provides for the coding of the identity of elements commonly found in organic compounds. Carbon and hydrogen are not coded directly but hydrocarbons are indicated when there are no punches in this column. In applying the code a punch is made in the proper position for each different element regardless of the number of such elements in the compound. The coding of less common elements is provided for in columns 56 and 57. A punch at the "y" position in column 32 must accompany any coding in these latter columns.

Column 33 codes the type and location of unsaturated carbon-to-carbon

Table 9-2A.

Row	Column 32 Elements	Column 34 Structure	Row	Column 36 Code Units	Column 38 Miscellaneous	Row
0	0	Acyclic	0	12 or more	Solid	0
1	N	Alicyclic	1	1	Liquid	1
2	S	Aromatic	2	2	Gas	2
3	F	Heterocyclic	3	3	Organo-metallic	3
4	Cl	Fused Alicyclic	4	4	Isotopic	4
5	Br, I	Fused Aromatic	5	5	Indeterminate	5
6	P, Bi	Fused Heterocyclic	6	6	Solution	6
7	As, Sb		7	7	Polymer	7
8	Si, Ge	3 or 4 Member Ring	8	8	Chelate	8
9	Sn, Pb	5 Member Ring	9	9	Hydrate	9
X	B, Al	6 Member Ring	X	10	KBr Plate	X
Y	Other	7 or more Member Ring	Y	11		Y

Row	Column 33 Unsaturation	Column 35 Rings—Chains	Row	Column 37 Substitutions	Column 39 Miscellaneous	Row
0	Ring	Rings	0	[mono]	cis	0
1	1	1	1	1 [1,2]	trans	1
2	2	2	2	2 [1,3]	spiro	2
3	3	3	3	3 [1,4]	dextrorotary	3
4	4	4	4	4 [1,2,3]	levorotary	4
5	5	5	5	5 [1,2,4]	symmetrical	5
6	6	6	6	6 [1,3,5]	unsymmetrical	6
7	7	7	7	7 [1,2,3,4]	vicinal	7
8	8	8	8	8 [1,2,4,5]	Salt	8
9	9	9	9	9 [1,2,3,5]	Inorganic Ester	9
X	$-\overset{\cdot}{\text{C}}=\overset{\cdot}{\text{C}}-$	10	X	10 [penta]		X
Y	$-\text{C}\equiv\text{C}-$	11 or more	Y	[hexa]	Inorganic	Y

bonds. In every case, except for aromatic unsaturation, the unsaturation is coded as to type, that is, double bond or triple bond or both, by punches at 33-x or 33-y, or both. Numbers in this column are used to indicate the location of the unsaturated bonds subject to the following rules:

- (1) If the unsaturation is located in a ring, then a 33-0 is required. When

Table 9-2B.

Row	Column 40 C - H	Column 42 O	Column 44 N	Row	Column 46 S	Column 48 N - O	Row
0	-CH ₃ methyl	-C(=O)OH	-C(=N)N<	0	-C(=S)S-	>NC(=O)O- -OC(=N)O-	0
1	-C ₂ H ₅ ethyl	-C(=O)O-	>NC(=N)N<	1	-C(=S)H	>NC(=O)N< -OC(=N)N<	1
2	-C ₃ H ₇ n-propyl	-C(=O)H	-CN	2	>C=S	-C(=O)N< -C(=N)O-	2
3	-C ₃ H ₇ iso-propyl	>C=O	-NC	3	-SC(=S)S-	-NCO	3
4	-C ₄ H ₉ n-butyl	-OC(=O)O-	-NH ₂	4	-SH	-OCN	4
5	-C ₄ H ₉ iso-butyl	C(OR) ₄	>NH	5	-S-	>NNO ₂	5
6	-C ₄ H ₉ sec-butyl	-C(OR) ₃	>N-	6	-SS-	>NNO	6
7	-C ₄ H ₉ ter-butyl	-OH	>C=N-	7	R ₃ S ⁺	-NN(=O)-	7
8	-C ₅ H ₁₁ n-pentyl	-O-	=NN= >NN<	8	=S=S	-ONO ₂	8
9	-C ₆ H ₅ phenyl	-OO-	-N=N- =N=N	9		-ON=O	9
X		Other	Other	X	Other	Other	X
Y		Heterocyclic	Heterocyclic	Y	Heterocyclic	Heterocyclic	Y

Row	Column 41 C - H	Column 43 O	Column 45 N	Row	Column 47 S	Column 49 N - O	Row
0	-CH=CH ₂ vinyl	R ₃ O ⁺	-N=N-N<	0		-NO ₂	0
1	>C=CH ₂ ter. methylene	-O ₃ -	-N≡N ⁺	1		-NO	1
2	=CHCH ₃ ethylidene	>C(-O-) ₂	R ₄ N ⁺	2		=NO- >NO-	2
3	-CH ₂ CH=CH ₂ allyl		-NH ₄	3		>N=O	3
4	-CH=CHCH ₃ propenyl		>NCN	4			4
5	-C(CH ₃)=CH ₂ isopropenyl		-N ₃	5			5
6	=C(CH ₃) ₂ isopropylidene			6			6
7	-C≡CH ethynyl			7			7
8	-C≡CCH ₃ 1-propynyl			8			8
9	-CH ₂ C≡CH 2-propynyl			9			9
X	Conjugated	Conjugated	Conjugated	X	Conjugated	Conjugated	X
Y				Y			Y

this punch is lacking, it is understood that unsaturation in a chain is being coded.

(2) Unsaturation at positions requiring numbers higher than 9, Greek letters, or primed numbers are not coded.

(3) The use of the position codes is restricted to compounds containing a

Table 9-2C.

Row	Column 50 N-S	Column 52 O-S	Row	Column 54 N-O-S	Column 56 Elements	Row
0	>NC(=S)N< -SC(=N)N<	-OC(=S)S- -SC(=O)S-	0	>NC(=S)O- -NC(=O)S-	Se, Te, Po	0
1	>NC(=S)S- -SC(=N)S-	-OC(=S)O- -OC(=O)S-	1	-C(=O)NS- -C(=S)NO-	Ga, In, Tl	1
2	-C(=S)N< -C(=N)S-	-C(=S)O- -C(=O)S-	2	-OC(=N)S-	Zn, Cd, Hg	2
3		-S(O ₂)O-	3	-S(=O)N<	Cu, Ag, Au	3
4	-SCN	-S(=O)O-	4	>NS(O ₂)N<	Fe, Co, Ni	4
5	-NCS	-S(=O)S- -S(=S)O-	5	>NS(O ₂)O-	Cr, Mo, W, U	5
6	>NSN<	S _x O ₆ (x = 2-6)	6	>NS(O ₂)-	V, Cb, Ta, Pa	6
7	=NS- -NS-	>SO ₂	7	>NS(=O)O-	Ti, Zr, Hf, Th	7
8	-N=S	>S=O	8	-NSO	Sc, Y, La, Ac	8
9	>S=N-	-OSO-	9		Ru, Rh, Pd, Os, Ir, Pt	9
X	Other	Other	X	Other	R.E.	X
Y	Heterocyclic	Heterocyclic	Y	Heterocyclic	Heterocyclic	Y

Row	Column 51 N-S	Column 53 O-S	Row	Column 55 N-O-S	Column 57 Elements	Row
0		-OS(O ₂)O-	0		Li	0
1		-OS(=O)O-	1		Na	1
2			2		K	2
3			3		Rb, Cs	3
4			4		Be	4
5			5		Mg	5
6			6		Ca	6
7			7		Sr, Ba	7
8			8			8
9			9			9
X	Conjugated	Conjugated	X	Conjugated	Conjugated	X
Y			Y		Other	Y

single chain, a single ring or a fused ring system where the Geneva System for chains or the Patterson Ring Index can be applied without ambiguity.

(4) Unsaturation in benzene rings, fused or otherwise, or in alicyclic rings as a result of fusion with aromatic rings is not coded here.

(5) Where both cyclic and chain systems are present in a single compound

TABLE 9-3. EXAMPLES OF THE TYPES OF COMPOUNDS CODED BY THE CODE UNITS IN THE CHEMICAL CLASSIFICATION CODE CHART

Following are examples of the types of compounds which the various code units in the chart may index. It is to be understood that these examples do not restrict the use of the code units in the indexing of other types of compounds in which they may appear.

42-0 acids	48-0 carbamyl compounds, carbamates
42-1 esters, salts, lactones anhydrides	48-1 ureido compounds
42-2 aldehydes	48-2 amides, imidic compounds, lactams
42-3 ketones	48-3 isocyanates
42-4 carbonates	48-4 cyanates
42-5 ortho carbonates	48-5 nitro amines
42-6 ortho carboxylates	48-6 nitroso amines
42-7 alcohols, phenols	48-7 azoxy compounds
42-8 ethers, oxy compounds	48-8 nitrates
42-9 peroxides	48-9 nitrites
43-0 oxonium compounds	49-0 nitro compounds
43-1 ozonides	49-1 nitroso compounds
43-2 acetals	49-2 isonitroso compounds, oximes
	49-3 amine oxides
44-0 amidines	
44-1 guanidines	50-0 thiourido compounds
44-2 nitrilo or cyano compounds	50-1 thiocarbamyl compounds
44-3 isonitrilo compounds	50-2 thioamides, thiomides
44-4 primary amines	50-3
44-5 secondary amines	50-4 thiocyano compounds
44-6 tertiary aminds	50-5 isothiocyano compounds
44-7 imines	50-6 diamino sulfides
44-8 hydrazones, hydrazines	50-7 sulfimes, sulfenamides
44-9 azo or diazo compounds	50-8 sulfamino and sulfinyl compounds
	50-9 sulfilimines
45-0 triazenes	
45-1 diazonium compounds	52-0 dithiocarbonates
45-2 quaternary ammonium compounds	52-1 thiocarbonates
45-3 ammonium compounds	52-2 thiolic, thionic compounds, carbodithioates
45-4 cyanamides	52-3 sulfonates
45-5 triazo compounds, azides	52-4 sulfinates
	52-5 thiosulfinates
46-0 thionoethiolic compounds, carbodithioates	52-6 thionates
46-1 thioaldehydes	52-7 sulfones
46-2 thiones, thioketones	52-8 sulfoxy compounds, sulfinyls
46-3 trithio carbonates	52-9 sulfenates
46-4 thiols	
46-5 sulfides	53-0 sulfates
46-6 disulfides, polysulfides	53-1 sulfites
46-7 sulfonium compounds	
46-8 perthio compounds	54-0 thiocarbamates
	54-1 carboxamido sulfides
	54-2
	54-3 sulfinamides
	54-4 sulfamides
	54-5 sulfamates
	54-6 sulfonyl amines, sulfonamides
	54-7 amino sulfinates
	54-8 sulfinyl amines
	54-9

and unsaturation is present in only one or the other, it is coded as to location.

(6) Where both cyclic and chain systems are present in a single compound and both contain unsaturation, the position code is applied to the largest ring or fused ring system.

Column 34 codes the major structural features of a compound and is largely concerned with the type and size of rings. The use of these codes in describing a molecular structure is governed by the following rules:

(1) An "acyclic" code is used whenever an open chain of two or more atoms other than hydrogen form a part or all of the molecule or whenever the molecule consists of only one atom other than hydrogen. Carbon atoms in rings are not counted as part of chains. Thus, methane, ammonia, phenylhydrazine, and ethylbenzene would require 34-0 punches, but toluene, phenol and aniline would not.

(2) Each individual type of ring present in a single molecule is coded by a single appropriate punch. Each member of a fused ring system is coded separately if different types are involved. All rings other than aromatic (benzene) and heterocyclic are considered alicyclic and only benzene rings are coded aromatic.

(3) No portion of any ring, except that involved in fusion, is coded more than once. Thus, multiple ring systems formed by bridging are individually coded but the enveloping ring is not.

(4) The size of aromatic rings is *not* coded with a 34-x punch.

Column 35 provides for coding the length of carbon chains or the number of rings in a compound. Use of the following rules will insure uniformity of coding:

(1) If there is only one ring, or if there are no rings in the compound, the length of the longest normal carbon-to-carbon chain is coded by an appropriate punch. One carbon atom is considered a "chain," but carbon atoms in rings are not to be counted as part of such chains.

(2) If there are two or more rings in the compound and aromatic rings are involved, both the total number of rings and the number of benzenoid rings are punched into column 35 together with a punch at 35-0. In any case, the total number of rings is punched in. Each ring in a fused ring, spiro or bridged system is counted separately. Rule 3 under column 34 also applies here.

Column 36 codes the total number and the number of different kinds of "code units" observed in the structure of the molecule. Both numbers should be coded into this column when there is a difference between the total number and the number of different kinds. The following rules assist in arriving at proper totals:

(1) Consider all code designations covered on the chart by columns 42

through 55 except codes for "heterocyclic" and "conjugated." Each designation made is a code unit.

(2) Consider all atoms other than C, H, N, O and S. Each other element counts as a code unit.

(3) For the total number of code units, count each unit as many times as it appears in the structure.

(4) For the number of different kinds of code units, count each type once.

Column 37 provides for locating the positions of substituent groups of "code units" in a limited number of cases. It is intended that this column provide a means of differentiating empirical isomers and is not rigorously applied in coding all compounds. One should not attempt to code substitutions in compounds where there is ambiguity as to just what is substituted on what. The following rules apply:

(1) Substitution positions requiring numbers higher than 10, or the use of Greek letters or primed numbers are not coded here.

(2) Except as provided in rule 4, below, use of the code is restricted to indicating substitution positions on a single carbon chain, or a single ring, or fused ring system where application of the Geneva System for chains and the patterson Ring Index for cyclic compounds can be made without ambiguity.

(3) In monocyclic compounds which also have acyclic components, code the location of substitutions on the ring.

(4) In polybenzenoid compounds not involving fusion, code designations within the brackets (on the chart) are used to indicate the degree and location of substitution on the several rings.

(5) The location of heteroatoms in heterocyclic rings are not to be made with this code.

Columns 38 and 39 code miscellaneous information about the compounds. For the most part they are self-explanatory, but the following interpretations should be made:

(1) Punches at 38-0, 1, 2 and 6 are used to indicate the physical state of the compound both at the time it is analyzed in the spectrometer and at room condition. Thus, 38-0, 6 indicates the material to be a solid at room temperature but was analyzed in solution.

(2) Punches at 39-5, 6 and 7 are not to be applied to coding trisubstitution on benzene or other cyclic compounds but rather to describe the arrangement of heteroatoms in heterocyclic rings such as the triazines where substitutions play no part in determining the use of the terms. Such application is not limited to rings containing one kind of heteroatom and the code may be applied to both five- and six-member rings.

Columns 40 and 41 code the smaller groups involving carbon and hydro-

gen only. The following rules apply:

(1) Code each unit that is observed in the structural formula of the compound being indexed.

(2) Use the largest code unit that will characterize a group and do not code the smaller parts of such a group. Thus, if a $\text{—C}_2\text{H}_5$ group is present, code a 40-1 but do not code a —CH_3 at 40-0.

(3) Under 41-x code all conjugated double bond systems involving carbon only, except purely aromatic conjugation. Do code conjugated carbon-carbon systems involving a benzene ring if there is at least one double bond outside of the ring or if two or more benzene rings form a part of a system.

(4) Always code the largest system, then do not code any of its parts.

Columns 42 through 55 provide for coding unit groups involving oxygen, nitrogen, or sulfur, singly or together, and with or without a single carbon atom. They are arranged in columns depending on the manner in which the atoms are involved. The following rules assist in the application of the code:

(1) Code each unit that is observed in the structural formula whether it is part of a ring or not. The only criterion is that the particular arrangement of atoms be present.

(2) Use the largest unit that will characterize a group and do not code any smaller parts. Thus, if the group NC(=O)O— is present, characterize it by a code of 48-0 and do not code 42-3, 42-8 or 44-6. Likewise, if the bonds of this code unit were satisfied with hydrogen atoms one would not use codes of 44-4 and 42-7 or 42-0.

(3) Code larger groups than appear in the chart, or those involving two or more carbon atoms, by using the *least number of largest code units* that are in the chart. Strict application of this rule is essential regardless of one's feeling for the chemistry or naming of compounds. In some cases this rule will necessitate the coding of an atom or two in each of two adjacent code units. Thus, for example, in $\text{CH}_2=\text{NNHC(=S)NHNH}_2$ one can observe the following code units:

C=N— , =NN , NC(=S)N , NN , C=S , NH , and —NH_2 . The problem is to include all of these structural arrangements in as few code units as possible. One begins by selecting a central carbon atom and observing the greatest number of heteroatoms attached to it. In the above example this process yields the

code unit NC(=S)N or 50-0. The other carbon atom calls for a code of

44-7 and all that remains are the two  groups which require a code of 44-8. Any other possible choice of code units would not involve the largest units and any further breakdown would not involve the least number of units. It will be noted that some of the nitrogen atoms were used twice in this process.

(4) Conjugated double bond systems involving the elements listed at the head of each column are coded with the appropriate "x" punch. A conjugated double bond system consists of a complete series of alternate double and single bonds which may extend through one or more benzene rings. One should code only the largest of any such system with the "x" that identifies the elements involved and not code any of its parts.

(5) The presence of heterocyclic rings involving elements listed at the head of the columns in the chart are coded by "y" punches in the appropriate columns. If two or more heteroatoms are involved in a single ring, use only the code that involves *all* of the atoms, except when one of the heteroatoms is other than N, O or S then a code of 56-y applies. Thus, a heterocyclic ring involving both O and S is coded as 52-y and codes of 42-y and 46-y are *not* used.

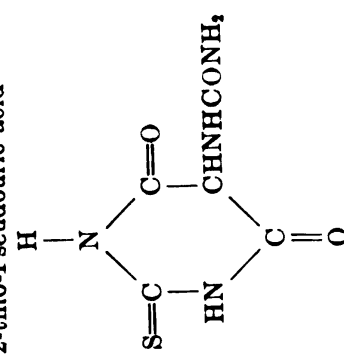
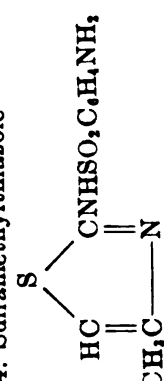
(6) Organic salts, including amine salts, etc., are to be coded in the unionized form and a code at 39-8 assigned to indicate a salt is involved. Organometallic codes are used only if there is a metal-to-carbon bond involved.

Columns 56 and 57 provide for coding the less common elements found in organic compounds. Provision is made to code conjugate systems and heterocyclic groups involving such elements and any element not listed in the chart can be coded at 57-y.

Application of the chemical classification code can best be facilitated by a study of examples. For these, the reader is referred to the many thousands of compounds that have been coded and punched into cards now commercially available. However, in order to provide an opportunity for one to obtain a degree of familiarity with the system with a minimum of effort, a number of examples are included in Table 9-4. Use of the coding system has spread to many areas far removed from infrared where it is desirable to characterize compounds according to structure and then to segregate or sort various classes and types of compounds independently from the names or chemistry associated with the compounds.

Sorting the cards to segregate compounds having structural features in common or for correlating structure and absorption bands involves merely an understanding of how the sorter operates, as described earlier, and knowing the code positions for the groups of interest. Both negative sorting and

TABLE 9-4. EXAMPLES OF CHEMICAL CLASSIFICATION

1. Parachlorophenol $\text{ClC}_6\text{H}_4\text{OH}$	32-0, 4 34-2 36-2	37-1, 4 38-0, 6 42-7
2. Isoprene $\text{CH}_2=\text{C}(\text{CH}_3)-\text{CH}=\text{CH}_2$	33-1, 3, x 34-0 35-4	37-2 38-1 41-0, 5, x
3. 2-thio-Pseudouric acid 	32-0, 1, 2 34-0, 3, x 35-1 36-3, 4	38-0 39-6 44-y 48-1, 2 50-0
4. Sulfamethylthiazole 	32-0, 1, 2 33-0, 2, x 34-0, 2, 3, 9 35-0, 1, 2 36-3 38-0 39-6	40-0 44-4 45-x 50-0, y 53-x 54-6

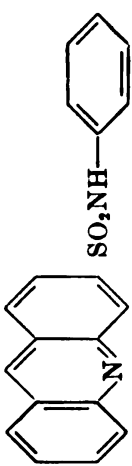
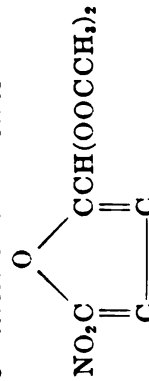
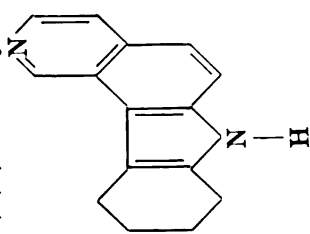
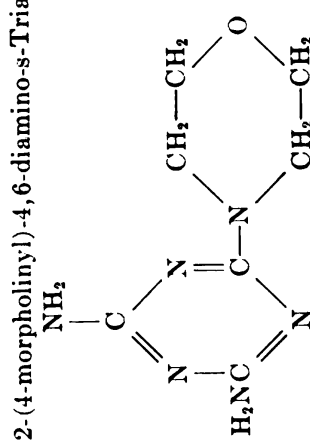
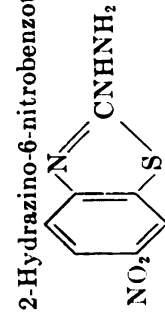
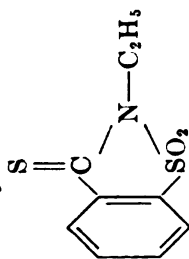
5. Potassium salt of gamma parachlorophenoxy crotylmercaptomethyl penicillin	$\text{CH}_3(\text{ClC}_6\text{H}_4\text{O})\text{C}=\text{CHCH}_2\text{SCH}_2\text{CNHCH}-\underset{\text{O}=\text{C}-\text{N}-\text{CHCOOK}}{\text{CH}-\text{S}-\text{C}(\text{CH}_3)_2}$	32-0, 1, 2, 4, y 33-2, x 34-0, 2, 6, 8, 9 35-0, 1, 3 36-6, 8 38-0 39-6, 8	40-0 42-1, 8 44-y 46-5 48-2 50-y 57-2
6. 3-Acridinesulfonanilide		32-0, 1, 2 34-0, 2, 5, 6, x 35-0, 3, 4 36-2 38-0	40-9 41-x 44-7, y 54-6 55-x
7. 3-Buten-2-ol, 1-bromo-3, 4-dichloro-	$\text{ClCH}=\underset{\text{Cl}}{\text{C}}-\underset{\text{Br}}{\text{CH}}\text{CH}_2\text{OH}$	32-0, 4, 5 33-3, x 34-0 35-4	36-3, 4 37-1, 2, 3, 4 38-1 42-7
8. Diethyl fluorophosphate	$(\text{C}_2\text{H}_5\text{O})_2\text{POF}$	32-0, 3, 6 34-0 35-2 36-4, 5	38-1 39-9 40-1 42-8, x
9. 5-Nitrofurfural diacetate		32-0, 1 33-0, 2, 4, x 34-0, 3, 9 35-2 36-4, 5 37-2, 5	38-0 40-0 42-1, 8, y 43-2 49-0, x

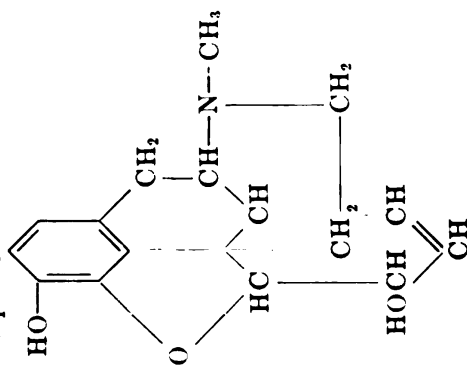
TABLE 9-4. *Continued*

10. 5,6,7,8 Tetrahydro-9H-pyrido (4, 3c) carbazole		32-1 33-0, x 34-4, 5, 6, 9, x 35-0, 1, 4 36-2 38-0 41-x 44-5, 7, y 45-x	
11. 2-(4-morpholinyl)-4,6-diamino-s-Triazine		32-0, 1 34-3, x 35-0, 2 36-2, 4 38-0 39-5 42-8	44-1, y 45-x 48-y
12. 2-Hydrazino-6-nitrobenzothiazole		32-0, 1, 2 34-0, 5, 6, 9 35-0, 1, 2 36-3 37-2, 6 38-0 39-6	44-8 49-0, x 50-0, y

13. 2-Ethyl thiosaccharin



14. Morphine



15. Chloroacetyl chloride



53-x
54-6

32-0, 1, 2
34-0, 5, 6, 9
35-0, 1, 2
36-2
37-1, 2, 3
38-0
39-7
40-1
50-2, y
32-0, 1
33-0, x
34-4, 5, 6, 9, x
35-0, 1, 5
36-3, 4
38-0
40-0
42-7, 8, y
44-6, y

36-2, 3
37-1, 2
38-1
42-3

32-0, 4
34-0
35-2

TABLE 9-4. *Concluded*

* The use of the terms symmetrical, unsymmetrical, and vicinal in connection with indicating the arrangement of heteroatoms will be useful here. Thus, ortho is vicinal, meta is unsymmetrical, and para is symmetrical. Heteroatoms may be the same or different.

32-0
34-3, x
36-1
38-1
39-6*
42-y
43-2

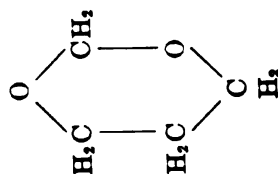
38-0
40-0
42-3, 7
43-x
44-6
48-2

39-2, 4, 8
42-0, 7, 8
44-4

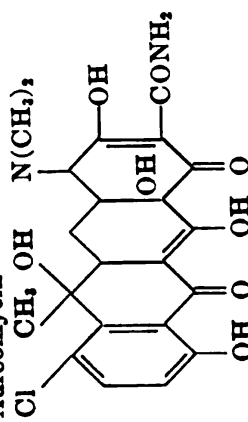
32-0, 1, 4
33-0, x
34-0, 4, 5, x
35-0, 1, 4
36-5, x

32-0, 1, 4, 5
34-0, 2
35-0, 2
36-6, 7
37-3, 9
38-0

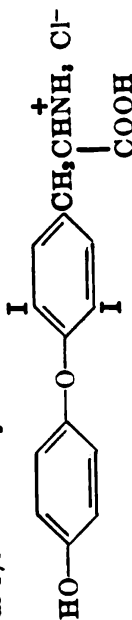
16. m-Dioxane



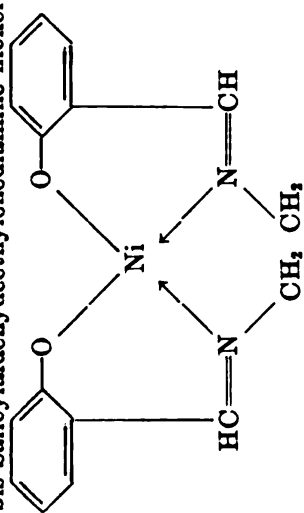
17. Aureomycin



18. dl-3,5-Diiodothyronine-HCl



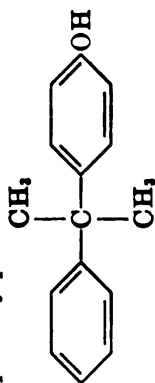
19. bis-Salicylaldehydeethylenediamine nickel



45-x
56-4, y

32-0, 1, y
34-5, 6, 9, x
35-0, 2, 5
36-3, 5
38-0, 8
39-7
42-8
44-7

20. p-Cumylphenol



38-0
40-0, 9
42-7

32-0
34-0, 2
35-0, 2
36-1
37-0, 3

21. Methyl ethyl ether



38-1
40-0, 1
42-8

32-0
34-0
35-2
36-1

22. n-Butyl acetate



38-1
40-0, 4
42-1

32-0
34-0
35-4
36-1

positive sorting are necessary, and the following illustrates the fundamental operations. Each sorting operation that can be performed narrows the character of the structure described by the segregated cards and sorting continues on residue decks until the precise structure desired is arrived at. Thus, if one wished merely to segregate all cards indexing compounds containing nitrogen, a *positive* sort at 32-1 would be adequate. Such a sort means that the number 1 switch is in the "on" position or pushed toward the circumference on the switch ring and all the rest are in the "off" position or pushed toward the center of the switch ring. (The large red switch always remains in the "on" position.) All cards dropping in the number 1 pocket will then index compounds containing at least nitrogen, and also all other elements. If it is desired to segregate cards indexing compounds containing nitrogen *only* in addition to carbon and hydrogen, a second sort on the residue deck is necessary. This is a *negative* sort at 32-1. In this case, the number 1 switch is in the "off" position and all the rest are in the "on" position. Now, all cards dropping in the reject pocket will index compounds having nitrogen *only* as the heteroatom. These are the two fundamental sorting techniques and the number of switches in the "on" or "off" positions, together with the column being sorted, will determine the types of structure, being either eliminated or segregated, as desired. To continue the sorting illustrations, one could take the nitrogen only cards and sort positively at 34-2 and 5, then take the cards from pockets 2 and 5 and sort negatively at 34-2, 5 and 0 to obtain all the cards indexing aromatic compounds only with nitrogen group substitutions either on the rings or side chains. The number of rings or the length of side chains could be regulated by sorts in column 35. The number and kinds of nitrogen groups could be limited by sorts in column 36 and, finally, the particular nitrogen groups could be segregated by sorts in columns 44 and 45. Each sorting operation is performed on the wanted residue from the previous sort. The sorting combinations are limitless and thus yield a very great variety of structure combinations. The A.S.T.M. instruction booklet¹⁴ contains additional examples of sorting operations.

A different chemical classification code has been provided for inorganic compounds. It is punched into the same areas of the card as the organic code but a special code punch at 39-y indicates that an inorganic compound is being coded. Therefore, cards indexing organic and inorganic compounds should be separated by the 39-y sort before sorting on chemical structure. In early issues of the infrared cards this inorganic code punch was located at 26-0. Otherwise, all infrared cards are identical and may be sorted together. The inorganic code is the same one used to index compounds whose x-ray diffraction powder patterns are in the A.S.T.M. files. The Elements Code, Table 9-5, provides for all elements with a direct punch code for each; the Radicals Code, Table 9-6, supplies additional

TABLE 9-5. ELEMENTS CODE

(43)*	32-0 Actinium—Ac	36-1 Nickel—Ni
	32-1 Aluminum—Al	36-2 Nitrogen—N
	32-2 Americium—Am	36-3 Osmium—Os
	32-3 Antimony—Sb	36-4 Oxygen—O
	32-4 Argon—A	[5] 36-5 Palladium—Pd
[1]†	32-5 Arsenic—As	36-6 Phosphorus—P
	32-6 Astatine—At	36-7 Platinum—Pt
	32-7 Barium—Ba	36-8 Plutonium—Pu
	32-8 Beryllium—Be	36-9 Polonium—Po
	32-9 Bismuth—Bi	36-x Potassium—K
	32-x Boron—B	36-y Praseodymium—Pr
	32-y Bromine—Br	
(44)	33-0 Cadmium—Cd	(48) 37-0 Prometheium—Pm
	33-1 Calcium—Ca	37-1 Proactinium—Pa
	33-2 Carbon—C	37-2 Radium—Ra
	33-3 Cerium—Ce	37-3 Rhenium—Re
	33-4 Cesium—Cs	37-4 Rhodium—Rh
[2]	33-5 Chlorine—Cl	[6] 37-5 Rubidium—Pb
	33-6 Chromium—Cr	37-6 Ruthenium—Ru
	33-7 Cobalt—Co	37-7 Samarium—Sm
	33-8 Columbium—Cb	37-8 Scandium—Sc
	33-9 Copper—Cu	37-9 Selenium—Se
	33-x Curium—Cm	37-x Silicon—Si
	33-y Dysprosium—Dy	37-y Silver—Ag
(45)	34-0 Erbium—Er	(49) 38-0 Sodium—Na
	34-1 Europium—Eu	38-1 Strontium—Sr
	34-2 Fluorine—F	38-2 Sulfur—S
	34-3 Francium—Fr	38-3 Tantalum—Ta
	34-4 Gadolinium—Gd	38-4 Technetium—Tn
[3]	34-5 Gallium—Ga	[7] 38-5 Tellurium—Te
	34-6 Germanium—Ge	38-6 Terbium—Tb
	34-7 Gold—Au	38-7 Thallium—Tl
	34-8 Hafnium—Hf	38-8 Thorium—Th
	34-9 Holmium—Ho	38-9 Thulium—Tm
	34-x Hydrogen—H	38-x Tin—Sn
	34-y Indium—In	38-y Titanium—Ti
(46)	35-0 Iodine—I	(50) 39-0 Tungsten—W
	35-1 Iridium—Ir	39-1 Uranium—U
	35-2 Iron—Fe	39-2 Vanadium—V
	35-3 Lanthanum—La	39-3 Ytterbium—Yb
	35-4 Lead—Pb	39-4 Yttrium—Yt
[4]	35-5 Lithium—Li	[8] 39-5 Zinc—Zn
	35-6 Lutecium—Lu	39-6 Zirconium—Zr
	35-7 Magnesium—Mg	39-7
	35-8 Manganese—Mn	39-8
	35-9 Mercury—Hg	39-9
	35-x Molybdenum—Mo	39-x
	35-y Neodymium—Nd	39-y Inorganic
(47)	36-0 Neptunium—Np	

^a Also: Helium—He; Krypton—Kr;
Neon—Ne; Radon—Rn; Xenon—Xe

* Numbers in parenthesis refer to columns used on the x-ray diffraction cards.

† Numbers in brackets refer to columns used on the formula-name cards.

TABLE 9-6. RADICALS

(51)* 40-0 aluminate	43-6 phosphate
40-1 ammonium	43-7 phosphide
40-2 antimonate	43-8 phosphite
40-3 antimonite	43-9 plumbate
40-4 arsenate	43-x plumbide
40-5 arsenide	43-y rhenate
40-6 arsenite	
40-7 bismuthate	(55) 44-0 selenate
40-8 borate	44-1 selenide
40-9 boride	44-2 selenite
40-x bromate	44-3 silicate
40-y bromide	44-4 silicide
	44-5 stannate
(52) 41-0 carbamate	44-6 stannide
41-1 carbide	44-7 sulphate
41-2 carbonate	44-8 sulphide
41-3 cerate	44-9 sulphite
41-4 chlorate	44-x tantalate
41-5 chloride	44-y telluride
41-6 chlorite	
41-7 chromate	(56) 45-0 tellurite
41-8 cyanamid	45-1 thionate
41-9 cyanate	45-2 titanate
41-x cyanide	45-3 thorate
41-y ferrate	45-4 tungstate
	45-5 uranate
(53) 42-0 ferrite	45-6 vanadate
42-1 fluoride	45-7 zincate
42-2 fulminate	45-8 zirconate
42-3 germanate	45-9 zirconyl
42-4 hafniate	45-x platinate
42-5 hexammine	45-y platinite
42-6 hydride	
42-7 hydroxide	(57) 46-0 chromite
42-8 iodate	46-1 gallate
42-9 iodide	46-2 palladite
42-x manganate	46-3
42-y molybdate	46-4
	46-5
(54) 43-0 nitrate	46-6
43-1 nitride	46-7
43-2 nitrite	46-8
43-3 osmate	46-9
43-4 oxide	46-x
43-5 pentammine	46-y

* Numbers in parenthesis refer to columns used on x-ray diffraction cards.

information. The use of suffixes and prefixes to further qualify these radicals has not been attempted; one will find, for example, that all phosphates, whether pyro-, ortho-, meta-, etc., will be coded by a punch at 6 in column 43. (See Tables 9-5 and 9-6). Finally, column 56 is reserved for miscellaneous items of the inorganic code according to the following:

56-0 Solid
56-1 Liquid
56-2 Gas
56-3 Solution
56-4 KBr Plate
56-5 Hydrate
56-6 Isotopic

(3) *Semi-empirical Formula*. Columns 58 through 62 provide space to record the number of C, N, O and S atoms in the compound being indexed. These values are punched directly into the columns as labeled on the card. Provision for indicating a larger number of atoms than 9 in columns 60, 61 and 62 is achieved by the use of overpunches. A "y" overpunch adds 10 to the value punched into the column, an "x" adds 20 and a "0" adds 30 to the number. A "0" punch alone would mean 30 atoms.

(4) *Melting or Boiling Point*. A melting or boiling point is punched into columns 63 through 65. Melting points are used when the material is a solid at and above 20°C and boiling points are at or near 760 mm pressure only. The following code identifies the number punched into these columns:

65-y—boiling point above 0°C
64-y—boiling point below 0°C
65-x—melting point.

(5) *Columns Reserved by A.S.T.M.*. Columns 29 through 31 have been set aside for possible future use as determined by A.S.T.M. Committee E-13 and should not be used for private codes because of possible future conflicts.

(6) *Columns Reserved for Private Use*. Columns 66 through 70 may be used by individuals as they see fit. At no time will Committee E-13 make code assignments to these columns. Such a section for private use will be found on all cards issued by the Committee.

(7) *Reference or File Number*. Columns 71 through 80 are used to describe the location of the standard or original spectral data from which the card was prepared. Since it is common practice to publish a number of infrared spectra on the same page in journals, it is not possible to refer to a specific spectrogram by means of an ordinary journal reference. Therefore, spectra abstracted by A.S.T.M.-sponsored groups are assigned serial numbers and a numerical index of these serial numbers together with the name of the

compound and the journal reference is published for users of the cards. Regular catalogs of infrared spectra bear individual serial numbers in which case the same serial numbers are used on the cards.

Letters punched into column 79 differentiate between the several collections of spectra according to the following code:

- 79-A—American Petroleum Institute
- 79-B—Users own file
- 79-C—Sadler Catalogue of Spectra
- 79-D—NRC-NBS File of Spectra
- 79-E—Spectra abstracted by A.S.T.M.
- 79-F—Documentation of Molecular Spectroscopy File.

Column 80 codes the type of data carried by the card and the following assignments have been made:

- 80-A—Infrared Absorption Data
- 80-B—X-Ray Diffraction Powder Data
- 80-C—Ultraviolet Absorption Data
- 80-D—Visible Absorption Data
- 80-E—Mass Spectral Data
- 80-F—Raman Data
- 80-G—Subject-Author Data
- 80-H—Near Infrared Data

Other assignments will be made as additional types of data are handled in the cards.

Workers at Dow Chemical Company¹⁵ have developed a method of identifying the components of a mixture by means of the infrared spectrogram, IBM cards and a collator. The stronger bands of standard spectra are coded by 9-row punches into columns representing the spectrum from 5 to 16 microns at intervals of from 0.1 to 0.5 microns. A selected list of structure groupings and classes are coded by "x" overpunches into these same columns. A collator can be wired to segregate in a single pass of the cards those representing spectra that (1) do not have bands at arbitrarily designated wavelength positions, (2) do have bands at other designated positions and (3) do have any number of designated grouping or class codes. This reduces the sorting time for the identification of mixtures. It was necessary to construct a special plugboard to simplify the rather time-consuming operation of wiring the standard IBM plugboard for each run.

Investigations carried on at Tennessee Eastman Company¹⁶ have shown

¹⁵ Baker, A. W., N. Wright, and A. Opler, "Automatic Infrared Punched-Card Identification of Mixtures," *Analytical Chemistry*, **25**, 1457-60, (1953).

¹⁶ Otis, M. V., "A Statistical Study of the Wyandotte-A.S.T.M. Punched Card Library of Infrared Absorption Spectra," presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Pittsburgh, Pennsylvania, March 3, 1955.

that the Wyandotte-A.S.T.M. cards can be handled in much the same way on the IBM Electronic Statistical Machine (101) so that a number of sorts, both positive and negative in any area of the card, can be made simultaneously with one pass of the cards through the machine to reduce sorting time. A unique feature of this method is the segregating of cards in several pockets depending upon whether they represent possible components of binary, ternary, quaternary or higher mixtures.

Near Infrared Absorption Spectroscopy

The recent development of specialized equipment has permitted a greatly increased use of the near infrared region of the spectrum for analytical purposes. Attempts to incorporate these new data into the existing code systems for the regular infrared region were generally unsatisfactory because a closer punching resolution for band positions is desirable. Altho a special hand-sorted card has not been proposed as yet, the Standard Data Subcommittee of ASTM Committee E-13 has approved an IBM card system to complement its existing systems and make better use of these new near infrared data. Since the purpose of the near infrared cards is essentially the same as that of the regular infrared cards previously described, the same codes apply to both in all respects except the coding of absorption data. Therefore, only those codes that apply to the punching of the spectral data are discussed here.

The near infrared spectral data indexing system for IBM cards is designed to handle absorption data in the region from 0.70 through 3.59 microns with a punching resolution of 0.01 micron. To be consistent with the other Wyandotte-ASTM Cards, band positions are coded in terms of wavelength in microns. At the head of columns 1 through 29 on the IBM card are printed numbers from 0.7 through 3.5 at intervals of 0.1 for each column. These numbers represent the whole number and tens values of the band positions and the hundreds values are punched into appropriate columns. Thus, the number printed at the head of column 20 is "2.6" so that a band position of 2.65 microns is coded by a single punch in column 20 at the "5" position. Determination of which bands to code follows the same general rules as prescribed for coding regular infrared spectra and sorting operations follow the same procedures previously described. It will be noted that the letter "H" in column 80 denotes near infrared data.

X-Ray Diffraction Powder Analysis

Qualitative analytical chemistry making use of x-ray diffraction powder data, essentially involves matching the diffraction pattern obtained from an unknown crystalline material with an identical pattern in a file of standard data obtained from known materials. Such a pattern consists of a number of diffraction lines of varying intensity separated by varying distances.

TABLE 9-7. HANAWALT GROUPS NUMBER CODE

d values	Group no.	d values	Group no.
Under 0.80	1	3.05-3.09	45
0.80-0.89	2	3.10-3.14	46
0.90-0.99	3	3.15-3.19	47
1.00-1.04	4	3.20-3.24	48
1.05-1.09	5	3.25-3.29	49
1.10-1.14	6	3.30-3.34	50
1.15-1.19	7	3.35-3.39	51
1.20-1.24	8	3.40-3.44	52
1.25-1.29	9	3.45-3.49	53
1.30-1.34	10	3.50-3.59	54
1.35-1.39	11	3.60-3.69	55
1.40-1.44	12	3.70-3.79	56
1.45-1.49	13	3.80-3.89	57
1.50-1.54	14	3.90-3.99	58
1.55-1.59	15	4.00-4.09	59
1.60-1.64	16	4.10-4.19	60
1.65-1.69	17	4.20-4.29	61
1.70-1.74	18	4.30-4.39	62
1.75-1.79	19	4.40-4.49	63
1.80-1.84	20	4.50-4.59	64
1.85-1.89	21	4.60-4.69	65
1.90-1.94	22	4.70-4.79	66
1.95-1.99	23	4.80-4.89	67
2.00-2.04	24	4.90-4.99	68
2.05-2.09	25	5.00-5.24	69
2.10-2.14	26	5.25-5.49	70
2.15-2.19	27	5.50-5.74	71
2.20-2.24	28	5.75-5.99	72
2.25-2.29	29	6.00-6.49	73
2.30-2.34	30	6.50-6.99	74
2.35-2.39	31	7.00-7.49	75
2.40-2.44	32	7.50-7.99	76
2.45-2.49	33	8.00-8.49	77
2.50-2.54	34	8.50-8.99	78
2.55-2.59	35	9.00-9.49	79
2.60-2.64	36	9.50-9.99	80
2.65-2.69	37	10.0-10.9	81
2.70-2.74	38	11.0-11.9	82
2.75-2.79	39	12.0-13.9	83
2.80-2.84	40	14.0-15.9	84
2.85-2.89	41	16.0-17.9	85
2.90-2.94	42	18.0-19.9	86
2.95-2.99	43	20.0-and over	87
3.00-3.04	44		

A.S.T.M. X-Ray Diffraction Data Cards.

TABLE 9-8. CODE FOR CHEMICAL COMPOSITION

Aluminum	Al	3-2	Neodymium	Nd	12-9
Americium	Am	13-12	Neptunium	Np	13-10
Antimony	Sb	9-7	Nickel	Ni	5-4
Arsenic	As	9-6	Nitrogen	N	9-4
Barium	Ba	2-12	Osmium	Os	6-12
Beryllium	Be	2-8	Oxygen	O	10-9
Bismuth	Bi	9-8	Palladium	Pd	6-11
Boron	B	3-1	Phosphorus	P	9-5
Bromine	Br	11-4	Platinum	Pt	6-1
Cadmium	Cd	4-10	Plutonium	Pu	13-11
Calcium	Ca	2-10	Polonium	Po	10-1
Carbon	C	7-2	Potassium	K	1-4
Cerium	Ce	12-7	Praseodymium	Pr	12-8
Cesium	Cs	1-6	Protactinium	Pa	13-8
Chlorine	Cl	11-3	Radium	Ra	2-13
Chromium	Cr	8-9	Rhenium	Re	8-3
Cobalt	Co	5-3	Rhodium	Rh	6-10
Columbium	Cb	8-11	Rubidium	Rb	1-5
Copper	Cu	5-6	Ruthenium	Ru	6-9
Curium	Cm	13-1	Samarium	Sm	12-11
Dysprosium	Dy	12-2	Scandium	Sc	3-4
Erbium	Er	12-3	Selenium	Se	10-12
Europium	Eu	12-13	Silicon	Si	7-3
Fluorine	F	11-2	Silver	Ag	5-7
Gadolinium	Gd	12-1	Sodium	Na	1-3
Gallium	Ga	3-6	Strontium	Sr	2-11
Germanium	Ge	4-12	Sulfur	S	10-11
Gold	Au	5-11	Tantalum	Ta	8-1
Hafnium	Hf	7-6	Tellurium	Te	10-13
Holmium	Ho	12-2	Terbium	Tb	12-1
Illinium	Il	12-10	Thallium	Tl	3-8
Indium	In	3-7	Thorium	Th	13-7
Iodine	I	11-5	Thulium	Tm	12-3
Iridium	Ir	6-13	Tin	Sn	4-13
Iron	Fe	5-2	Titanium	Ti	8-13
Lanthanum	La	12-6	Tungsten	W	8-2
Lead	Pb	4-1	Uranium	U	13-9
Lithium	Li	1-2	Vanadium	V	8-7
Lutecium	Lu	12-5	Ytterbium	Yb	12-4
Magnesium	Mg	2-9	Yttrium	Y	3-5
Manganese	Mn	8-10	Zinc	Zn	4-9
Mercury	Hg	4-11	Zirconium	Zr	7-5
Molybdenum	Mo	8-12			

into the second and third line fields with the 7-4-2-1 code. This system makes it possible to use any one of the three strongest lines to determine a Hanawalt Group in which the sought-for card must be, and then to sort notches on the second and third lines to isolate the desired cards. The cards are filed according to the groups so that the first sort results from the act of removing the cards from the file. A code designation of the elements in the compound (Table 9-8) is notched into the chemical composition field of the card and may be used as a cross-sort at any time such information is known. Considerable space is left in the edges of the card for notching additional information such as melting point, indices of refraction, etc., but this can be done at the discretion of each individual user. On the face of the card are printed the complete crystallographic and x-ray powder data of the compound together with name, formula, purity, source and other pertinent notes. A system for tabulating detailed crystallographic data in Key-sort cards has been proposed by workers at Armour Research Foundation.²⁰ These cards contain all of the data normally published in the Crystallographic Data series in "Analytical Chemistry" and provides an excellent method of searching the data for identification purposes.

Machine-Sorted (IBM) Cards. A method of applying International Business Machines Company cards and equipment to the problems of sorting and correlating x-ray diffraction powder data was developed at Wyandotte Chemicals Corporation.²¹ Through the cooperation of members of the Joint Committee on Chemical Analysis by Powder Diffraction Methods this system was accepted by the A.S.T.M. and punched cards indexing all of the powder diffraction data are available. Supplements of the IBM cards are offered with each set of the x-ray data as they are released by A.S.T.M. The Wyandotte-A.S.T.M. x-ray diffraction card, codes and sorting methods are modifications of the infrared absorption indexing system previously described and much of the descriptive material pertaining to the handling of IBM cards presented in the infrared section is equally applicable here. The following description includes the necessary codes that enable one to prepare and use the cards. Additional details and examples may be obtained from the instruction booklet distributed by A.S.T.M.²²

The card for indexing x-ray diffraction data (Figure 9-5) is divided into

²⁰ McCrone, W. C., "Punched-Card System for Tabulating of Crystallographic Data," *Analytical Chemistry*, **28**, 972-5 (1956).

²¹ Kuentzel, L. E., "The Use of Hollerith Punched Cards for Indexing X-Ray Diffraction Powder Data," presented before the American Crystallographic Association, Chicago, Illinois, October 24, 1951.

²² Kuentzel, L. E., "Codes and Instructions for Wyandotte-A.S.T.M. Punched Cards Indexing X-Ray Diffraction Powder Data," Wyandotte Chemicals Corporation, Wyandotte, Michigan (1951) and IBM Technical Newsletter No. 4, International Business Machines Corporation, New York City (1953).

[illegible]

the following areas for coding purposes:

- (1) Diffraction Line Spacings—columns 1 through 35
- (2) Hanawalt Group Code—columns 36 and 37
- (3) Chemical Classification—columns 43 through 62
- (4) Melting Point—columns 63 through 65
- (5) Reserved for A.S.T.M.—columns 38 through 42
- (6) Reserved for Private Use—columns 66 through 70
- (7) Reference or Serial Number—columns 71 through 80

All available data about a given compound are coded into one card and three copies of the card, identical except for the group code number, are included in the file. Punches at 1, 2 or 3 in column 27 indicate which of the three strongest lines was used in determining the group code for the particular card.

(1) *Diffraction Line Spacings:* All “d” lines of the diffraction pattern having an intensity one-tenth or more of the intensity of the strongest line are punched into the cards. This is achieved by punching the final digit into the column headed by the number supplying the rest of the digits. The column headings are printed on the card but Table 9-9 gives the column heading code.

Values below 1.00 Å are rounded off to the nearest tenth angstrom and punched directly into column 1. From 1.00 through 3.40 angstrom units the punching resolution is 0.01 unit and the value of the hundreds digit is punched into the column having the proper heading. Thus, a value of 1.98 Å is coded by a single punch at the 8 position in column 11 which is headed by the number 1.9. Beginning with column 27 and through 33 the punching resolution is again 0.1 Å. Thus, a value of 23.7 Å is coded as a 4 punch in column 35. All values greater than 29 Å are coded as a 9 punch in column 35. Use of the selector switch on the sorter, as described in the discussion on the infrared system, enables one to segregate all cards bearing any given

TABLE 9-9. LINE SPACING CODE

Column	Line	Column	Line
1	0	19	2.7
2	1.0	20	2.8
3	1.1	21	2.9
4	1.2	22	3.0
5	1.3	23	3.1
6	1.4	24	3.2
7	1.5	25	3.3
8	1.6	26	3.4
9	1.7	27	3
10	1.8	28	4
11	1.9	29	5
12	2.0	30	6
13	2.1	21	7
14	2.2	32	8
15	2.3	33	9
16	2.4	34	10
17	2.5	35	20
18	2.6		

line combination. Also, as with the infrared system, one can sort over a narrow range of values simultaneously if the exact value of the line is in doubt. With these cards, since a great many lines are coded in each, sorting operations are not confined to the three strongest which may be less characteristic than other lines in the pattern.

(2) *Hanawalt Group Code.* Use is made of the three-card-per-compound and group code system for filing the punched cards in Hanawalt groups, as described previously for hand-sorted x-ray diffraction cards. This enables one to use any one of the three strongest lines to determine a group in which the wanted card must be and then to withdraw only that group from the files for sorting operations. For reason of uniformity, the same group code as suggested by Dr. Hanawalt and used on the Keysort cards previously described is used on the IBM cards (Table 9-7). The code group numbers are punched directly into columns 36 and 37. This enables the cards to be sorted into Hanawalt groups by machine for filing purposes.

(3) *Chemical Classification.* The identity of all elements, inorganic radicals, type of compound and other pertinent data are coded into this section. This enables one to make card-eliminating sorts in this section when such information is available. This has proved to be a particularly powerful tool whenever it can be used. The codes involved are direct and the methods employed are similar to those described for structure sorts in the section on infrared. The Elements Code for section A on the card and the Radicals Code for section B are the same as used for inorganic compounds on the infrared card except that different columns are involved. The column values for use on the x-ray cards will be found in parentheses in Tables 9-5

and 9-6. Codes for sections C and D on the x-ray card are given below:

Section C, Organic

59-0 Saturated aliphatic
59-1 Unsaturated aliphatic
59-2 Saturated monocyclic
59-3 Unsaturated monocyclic
59-4 Saturated polycyclic
59-5 Unsaturated polycyclic
59-6 Benzo aromatic
59-7 Polybenzo aromatic
59-8 Fused ring aromatic
59-9 Heterocyclic
59-x Unassigned
59-y Unassigned

Section D, Miscellaneous

61-0 Hydrated
61-1 Inorganic
61-2 Organic
61-3 Metal organic
61-4 Unassigned

(4) *Melting Point.* Columns 63, 64 and 65 provide for recording melting points. All melting points higher than 999°C are punched as 999. If the value punched into the card is a negative number, the fact is indicated by an additional "x" overpunch in column 65. Melting points have not been generally available together with the diffraction data, so this feature of the cards as currently distributed is far from complete.

(5) and (6) *Reserved Areas.* Columns 38 through 42 are reserved for future use by A.S.T.M., and columns 66 through 70 are reserved for private use by individual laboratories.

(7) *Reference or Serial Number.* This section of the card provides space to record a reference to the original data from which the card was prepared. Provision is made for either a journal or book reference or to the A.S.T.M. data card serial number. So far, only the latter type of reference has been used on published punched cards. The serial number is punched directly into columns 75 through 78, the set designation is coded into column 79 where "A" means Set 1, "B" means Set 2, "C" means Set 3, etc. On the A.S.T.M. x-ray data cards, the sets are designated by a digit in front of the serial number. Thus, data identified by serial number 3-0725 on the A.S.T.M. x-ray data card is punched into an IBM card bearing the number 725C. The code used for punches into column 80 is the same for all types of Wyandotte-A.S.T.M. cards and the assignments are listed earlier in the infrared section of the chapter.

Normally, the x-ray IBM cards are filed according to Hanawalt Group

numbers so that any given group or groups may be removed directly from the files for sorting operations. Any one of the three strongest lines may be used to determine group selected. The first actual sort on the cards will depend upon what other information is available. If the pattern alone must form the basis of the search, the most uncommon line should be used for the first sort in order to eliminate as many cards as possible. This may be the innermost line or an unusually strong line in the short-spacing section of the pattern. The operations are carried out in exactly the same way as described for the sorting of infrared absorption bands earlier in this chapter. Additional sorts on each residue deck continue to eliminate unwanted cards. If a reliable sort can be made on a metal element, the search can be narrowed very rapidly. In any event, the order of the sorting operations does not alter the final results and because of the relatively small number of cards in the Hanawalt Groups, the operations require but a very few minutes. The serial number on the final card or cards refers one directly to the complete data and names in the A.S.T.M. powder diffraction file for a final comparison with the unknown before the identification is accepted. A recent paper by Beukelman^{22A} describes efficient sorting operations for effective use of these cards.

Ultraviolet Absorption Spectroscopy

The application of ultraviolet absorption data to qualitative analytical determinations involves much that is similar to the use of infrared absorption data. The ultraviolet spectrogram is a fingerprint of the compound that produced it, although it is usually somewhat less detailed than the normal infrared spectrogram. In such matters as the coding and sorting of absorption band positions, the identity of important elements, a classification of the chemical structure and melting or boiling points, both methods are quite similar. Moreover, the fact that ultraviolet absorption data can be recorded and published in a number of different ways provides the same problems of matching unknown data with published data. However, in general, the ultraviolet spectra offer less detail for comparison purposes and the effects of solvents are more pronounced. This is reflected in the design of cards to index ultraviolet data for sorting purposes. It becomes necessary to include details on the intensity of critical absorption bands and to identify the solvent used. The very large mass of published ultraviolet spectral data has made the use of punched card indexing methods for universal searches a necessity.

Hand-Sorted Cards. At present there is no widely accepted and used notched card system for handling ultraviolet absorption data. The Na-

^{22A} Beukelman, T. E., "Efficient Use of IBM File of ASTM Powder X-Ray Diffraction Data," *Analytical Chemistry*, 29, 1269-72 (1957).

tional Research Council Committee on Spectral Absorption Data⁷ is working on Keysort card to be a companion to the infrared card currently being distributed. Present status of the card provides for coding principal absorption bands from 200 to 400 millimicrons, a field for band-no band coding in 50 millimicron steps, chemical structure, solvents and specific absorbence values of the strongest bands. Printed on the card will be the name and formulas of the compound, physical state, solvent concentration, cell thickness, source and purity of compound and contributing laboratory. Meanwhile, another proposal for such a card has been published²³ and general interest is rapidly increasing. However, since none of the cards or systems is being offered commercially at the present, the codes are not in sufficiently final form to be included here.

Machine-Sorted (IBM) Cards. A logical extension of the methods used in handling infrared absorption data into the field of ultraviolet and visible absorption spectroscopy was made at Wyandotte Chemicals Corporation.²⁴ This IBM system, with modifications contributed by members of A.S.T.M. Committee E-13, has been adopted by the American Society for Testing Materials as a standard method of indexing and sorting such data¹⁴ and large decks of punched and printed cards are available from the Society. The collecting and abstracting of ultraviolet absorption data for these cards are in the hands of the same A.S.T.M. committees and the cards are being prepared by the same group at the National Bureau of Standards as described earlier in this chapter. The sorting techniques outlined for use with the infrared Wyandotte-A.S.T.M. cards are applicable to the ultraviolet cards to be described. The following description includes all of the codes necessary for the proper use of the cards in searching published ultraviolet absorption data for qualitative analytical purposes.

The Wyandotte-A.S.T.M. card indexing ultraviolet absorption data (see Figure 9-6) is divided into the following areas for coding purposes:

- (1) Ultraviolet Absorptions—columns 1 through 11
- (2) Number of Peaks—columns 12 and 13
- (3) Intensity of Peaks—columns 14 through 17
- (4) Solvents and pH—columns 29 through 31
- (5) Chemical Classification—columns 32 through 57
- (6) Semi-empirical Formula—columns 58 through 62
- (7) Melting or Boiling Point—columns 63 through 65
- (8) Reserved by A.S.T.M.—columns 18 through 28
- (9) Reserved for Private Use—columns 66 through 70
- (10) Reference or Serial Number—columns 71 through 80

²³ Kendall, C. E., "Indexing of Data on Ultraviolet Absorption Spectroscopy," *Applied Spectroscopy*, **9**, 158-165 (1955).

²⁴ Kuentzel, L. E., "The Indexing and Sorting on IBM Equipment of Infrared, Ultraviolet, Mass and Other Standard Data," paper presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Pittsburgh, Pennsylvania, March 6, 1952.

Figure 9-6. Wyandotte-A.S.T.M. Ultraviolet Data Card.

All of the data from one compound are punched into one card. Codes for items (5), (6), (7) and part of (10) are identical to those used for the same columns on the infrared card and since they are described in detail in an earlier section of this chapter, they will not be repeated here.

(1) *Ultraviolet Absorptions.* The coding of the positions of ultraviolet absorption bands or peaks is done in terms of wavelength in millimicrons. The coding resolution is $2\text{ m}\mu$, that is, any number of peaks $2\text{ m}\mu$ or more apart may be coded individually and peaks closer than this are coded as one. The wavelength interval covered by each column is printed at the head of the column. Thus, a peak value of $244\text{ m}\mu$ would be coded by a single punch at the number 2 position in column 3 which is headed by the number 240. Each successive digit in the column represents an increment of $2\text{ m}\mu$ over the preceding one and the 0 punch value is that which is printed at the head of the column. Again, a value of $338\text{ m}\mu$ is coded by a punch of 9 in column 7 which has the value 320 at its head. To indicate the range of the spectra data covered by the particular spectrogram being coded, an "x" overpunch is placed in each column for which no data are available. A "y" overpunch in the same column where the last measurements on a terminal absorption are recorded indicates that there is a possible band just outside the range covered by the published data. Finally, the general position of the longest wavelength band in the spectrum is indicated with an "x" overpunch in the same column with the punch designation of the peak position. Such an "x" overpunch need not be confused with the "no data" overpunches because in the latter case no other punches appear in the same column.

Sorting the peak positions with the ultraviolet indexing cards is essentially the same as the methods used in handling the cards indexing infrared spectral data. Both positive and negative sorting approaches are feasible. Use of the longest wavelength band "x" overpunch is an effective tool. Thus, if there is an unknown with its longest wavelength band at 314 mil-

TABLE 9-10

Column	Punch	Range (m μ)	Peaks
12	0	200 to 250	None
12	1	200 to 250	One peak
12	2	200 to 250	Two peaks
12	3	200 to 250	Three or more
12	5	250 to 300	None
12	6	250 to 300	One peak
12	7	250 to 300	Two peaks
12	8	250 to 300	Three or more
13	0	300 to 350	None
13	1	300 to 350	One peak
13	2	300 to 350	Two peaks
13	3	300 to 350	Three or more
13	5	350 to 400	None
13	6	350 to 400	One peaks
13	7	350 to 400	Two peaks
13	8	350 to 400	Three or more

limicrons, sorts can be made to eliminate all spectra that do not have their longest band in the 300 to 318 m μ range (column 6) by sorting on the "x" overpunch.

(2) *Number of Peaks.* The number of peaks in each 50 millimicron interval of the spectrum are indicated by punches in columns 12 and 13, according to the given code in Table 9-10.

Thus, a spectrogram which exhibits peaks at 225, 314, 322, 345 and 375 millimicrons would be coded by punches at 12-1 and 13-3,6. A special punch at "y" in column 12 is used to indicate that the data coded into the ultraviolet absorption section of the card was published in tabular form. In such cases no "x" overpunches are used except to locate the position of the longest wavelength band.

(3) *Intensity of Peaks.* The intensity, in terms of absorbence for a solution of 1 gram per liter in a 1-cm cell (the absorptivity), of the strongest peak in each 50-m μ interval is punched into columns 14 through 17 by means of the code given in Table 9-11. These data are included so that one can make use of the differences in peak intensities to separate spectra with peaks located at the same wavelength positions but exhibiting differences in intensity.

(4) *Solvents and pH.* Because of the influence of solvent upon the shape of ultraviolet spectra, it was deemed advisable to provide a method of segregating spectra obtained from compounds in different solutions. In Table 9-12 is a direct code for a number of solvents commonly used. The solvents are arranged in order of frequency of use as revealed by a study of a large number of published spectra.

In order to provide a finer breakdown of water solutions based upon pH,

TABLE 9-11

Column	Range (m μ)	Punch	Intensity
14	200 to 250	0	0 to 1
15	250 to 300	1	1 to 3
16	300 to 350	2	3 to 10
17	350 to 400	3	10 to 30
		4	30 to 50
		5	50 to 75
		6	75 to 100
		7	100 to 200
		8	200 to 300
		9	300 to 1000
		x	Over 1000

TABLE 9-12

Column	Punch	Solvent
29	0	Aliphatic hydrocarbon; isooctane, cyclohexane, etc.
29	1	95% Ethanol
29	2	Absolute ethanol
29	3	Absolute methanol
29	4	1 Normal or stronger acid
29	5	0.1 Normal acid solution (HCl)
29	6	Water; pH 5 to pH 9
29	7	0.1 Normal base solution (NaOH)
29	8	1 Normal or stronger base
29	9	Dioxane and water mixtures
29	x	CHCl ₃ , CCl ₄ , SC ₂
29	y	Aromatic hydrocarbons; benzene, toluene, etc.
30	0	Glacial acetic, conc. H ₂ SO ₄ , etc.
30	1	Ethers
30	2	Ketones, esters
30	3	Pyridine and other basic solvents
30	4	Dimethyl formamide, dimethyl acetamide
30	5	Other
30	6	Solvent unknown, not reported, etc.
30	7	No solvent—vapor, film, liquid, gas, etc.

a direct code is supplied. This information, when available, is punched into column 31 according to the following code:

Punch	pH	Punch	pH
0	below 1	5	7
1	1 or 2	6	8
2	3 or 4	7	9
3	5	8	10 or 11
4	6	9	above 11

(8) and (9). *Reserved Sections.* Columns 18 through 28 are reserved for future use by A.S.T.M. action, and columns 66 through 70 are reserved for private use as on the infrared indexing card.

(10) *Reference or File Number*. This section of the card is used to identify the source of the data coded into the card. As with the infrared indexing cards, it has not been feasible to use journal references directly so serial numbers are assigned and lists of serial numbers, compound names and journal references are issued by A.S.T.M. to users of the cards. Where publishers have already assigned serial numbers to the spectra, as is done by the American Petroleum Institute Research Project 44, the same serial numbers are used on the cards. Following are the code assignments to column 79 for indicating the source of ultraviolet data:

Code	Source
A	American Petroleum Institute Research Project 44
B	User's own file of spectra
C	Spectra issued by the NRC-NBS Committee
D	Spectra abstracted by A.S.T.M.-sponsored groups

Others will be added as needed. The code used in Column 80 is the same as that used on the infrared card and may be obtained by referring to the appropriate section of this chapter.

Visible Absorption Spectroscopy

The application of visible absorption data to qualitative analytical determinations follows essentially the same pattern as developed for ultraviolet absorption data. There is only the added feature that the compounds and solutions involved usually have color, and it is convenient to have a means of indicating this fact. Because of the similarity of the systems involving infrared, ultraviolet and visible absorption data for qualitative work, a general discussion will not be given here and the reader is referred to the appropriate previous discussion in this chapter for background information.

Hand-Sorted Cards. At present, there is no generally accepted and used punched card system for handling visible absorption data. The National Research Council Committee on Spectral Absorption Data⁷ has plans for a Keysort card covering the range of 400 to 800 m μ . It will be quite similar to the one nearly completed for the ultraviolet region which was previously described. The only major publication on the subject²³ combines ultraviolet and visible data in one card at some sacrifice in coding resolution.

Machine-Sorted (IBM) Cards. A system for handling visible absorption data in IBM cards was proposed by workers at Wyandotte Chemicals Corporation.²⁴ This system, with modifications contributed by members of A.S.T.M. Committee E-13, has been adopted by the American Society for Testing Materials as a standard method of indexing and sorting such data¹⁴, and decks of punched and printed cards are available from the Society. The collection and editing of the published data and the prepara-

TABLE 9-13

Column	Punch	Range ($m\mu$)	Peaks
11	0	350 to 450	None
11	1	350 to 450	One peak
11	2	350 to 450	Two peaks
11	3	350 to 450	Three or more
11	5	450 to 550	None
11	6	450 to 550	One peak
11	7	450 to 550	Two peaks
11	8	450 to 550	Three or more
12	0	550 to 650	None
12	1	550 to 650	One peak
12	2	550 to 650	Two peaks
12	3	550 to 650	Three or more
12	5	650 to 750	None
12	6	650 to 750	One peak
12	7	650 to 750	Two peaks
12	8	650 to 750	Three or more
13	0	750 to 850	None
13	1	750 to 850	One peak
13	2	750 to 850	Two peaks
13	3	750 to 850	Three or more

in 5 or 0 before coding. Wavelength intervals covered by each column are printed at the head of the column. The "0" punch value is that printed at the head of the column and each successive digit in the column represents an increment of 5 $m\mu$. Thus, a value of 560 $m\mu$ is coded by a punch at the 2 position in column 5 which is headed by the number 550. Overpunch codes used in this section are the same as used on the ultraviolet card.

(2) *Number of Peaks.* The number of peaks in each 100 $m\mu$ interval of the spectrum are indicated by punches in columns 11, 12 and 13 according to the direct code given in Table 9-13. Thus, a spectrogram that exhibited peaks at 375, 560, 575, 645 and 770 $m\mu$ would require code punches at 11-1,5, 12-3 and 13-1.

(3) *Intensity of Peaks.* The intensity, in terms of absorbence for a solution of 1 gram per liter in a 1-cm cell (the absorptivity), of the strongest peak in each 100- $m\mu$ interval is punched into columns 14 through 18. The intensity code used is the same as provided for coding the intensity of ultraviolet absorption peaks. (See Table 9-11) The range code for the visible card follows:

Column	Range ($m\mu$)
14	350 to 450
15	450 to 550
16	550 to 650
17	650 to 750
18	750 to 850

Thus, a peak at $625\text{ m}\mu$ having an intensity of 25 would be coded as a 3 punch in column 16.

(11) *Reference or File Number.* As with the other A.S.T.M.-sponsored cards, the serial number of the spectrogram is punched into columns 73 and 78 and the type of data coded into the card is indicated by punches in column 80. The source of the visible data are coded into column 79 as follows:

- 79-A—(To be assigned)
- 79-B—User's own file of spectra
- 79-C—Spectra issued by the NRC-NBS Committee
- 79-D—Spectra abstracted by A.S.T.M.-sponsored groups.

The column 80 code is given in the section on infrared.

Mass Spectrometry

The mass spectrum of a compound provides a unique set of data which can be used for qualitative analysis. In part, such an analytical operation involves a comparison of mass spectral data obtained from the unknown material with that obtained from known standard materials. Such a mass spectrum is rather complex and is usually represented by the actual trace from a recorder, a schematic drawing or a tabulation of the various mass-charge ratios and relative intensities. A listing of certain other operational factors essential to the production of comparable data usually accompanies such mass spectra. A large library of mass spectra and a good means of sorting and indexing it are essential to effective and efficient qualitative analysis. The ever-growing accumulation of mass spectral data available from the American Petroleum Institute Research Project 44 provides such a library and many laboratories have made notched card files of this and other data to facilitate the necessary matching operations.

Hand-Sorted Cards. Two systems employing Keysort cards to facilitate sorting of mass spectral data have attracted considerable attention. They are sponsored by two manufacturers of mass spectrographs, namely, Consolidated Electrodynamics Corporation and General Electric Company.* Since the C.E.C. system has been incorporated into cards that are commercially available²⁵, it will be described in some detail.

The Consolidated Electrodynamics Corporation card (Figure 9-8) pro-

* Thanks are due The Consolidated Electrodynamics Corporation, Pasadena, California, for permission to reproduce in this chapter its copyrighted mass spectrum card.

²⁵ "Keysort File of Mass Spectra," Consolidated Electrodynamics Corporation, Pasadena 8, California.

ION MASS		ELEMENTAL		BOILING POINT (BY °C)		MOLECULAR WEIGHT	
CARD No. 529		MASS SPECTRUM				COPYRIGHT BY	
AMERICAN PETROLEUM INSTITUTE, PROJECT 44, CATALOG OF MASS SPECTRAL DATA						CONSOLIDATED ENGINEERING CORPORATION	
OTHER:						SERIAL NO. 624	
LABORATORY						DATE OF MEASUREMENT	
The Atlantic Refining Co., Philadelphia, Pa.						March 31, 1951	
MASS SPECTROMETER				COMPOUND			
MODEL Consolidated 21-101 (modified)				NAME N-n-Butylpyrrole			
ELECTRON CURRENT:				SOURCE			
(m/e) VOLTS (m/e) VOLTS				Eastman Kodak Co.			
ION ACCELERATING VOLTAGES 57 1880 200 530				PURITY MOLE			
ION SOURCE TEMPERATURE °C				WEIGHT 123.19 MOLECULAR FORMULA C ₈ H ₁₃ N			
METASTABLE SUPPRESSOR NONE [] NORMAL [] HIGH []				SEMI-STRUCTURAL FORMULA C ₈ H ₁₃ NHCH ₂ CH ₂ CH ₂ CH ₃ BOILING POINT °C 91.2			
BASIS OF PRESSURE MEASUREMENT PVT Measurement				n-BUTANE PATTERN AND SENSITIVITY			
ADDITIONAL INFORMATION				29 44.1 58 10.4 43 100 SENS ₄₃ 50.00 d.u./μ			
				SENSITIVITY *m/e 81 59.38 d.u./μ			
PATTERN OR RELATIVE INTENSITIES AT 70 IONIZING VOLTS							
m/e	PATTERN	m/e	PATTERN	m/e	PATTERN	m/e	PATTERN
37	1.41	56	.68	92	.50		
38	3.93	57	.91	93	1.57		
39	17.6	65	1.23	94	4.59		
40	6.04	66	2.97	95	1.80		
41	14.7	67	12.4	106	.72		
42	2.06	68	5.62	108	2.18		
43	.92	77	.61	122	3.95		
44	2.57	78	2.29	123	35.0		
50	2.22	79	1.85	124	3.18		
51	4.66	80	51.2				
52	3.46	81	100.00				
53	15.3	82	4.80				
54	4.56						
55	2.88						

CONSOLIDATED ENGINEERING CORP. CHEMICAL INSTRUMENT DIVISION - FORM NO. 427

Figure 9-8. Consolidated Electrodynamics Corp. Mass Spectrum Card.

vided for notching the following information:

- (1) Molecular Weight
- (2) Boiling Point
- (3) Elements
- (4) Ion Mass of Peaks

All the data available concerning one compound are punched into or printed on a single card. This includes, in addition to the data listed above, the name and formula of the compound, source and purity, type of instrument, accelerating voltages, serial numbers and other pertinent data.

(1) and (2) *Molecular Weight and Boiling Point*. These values are notched into the designated areas of the card by means of the familiar, 1, 2, 4, 7 system. Provision for molecular weights to 999 is made and boiling points are punched in at 10°C intervals with only the tens and hundreds digits being notched. A special position is reserved to indicate that the number punched has a negative value.

(3) *Elements*. The identity and a rough indication of the number of the common elements of organic chemistry are coded into one section of the

card according to the following schedule of shallow and deep punches:

Element	Number of Atoms	
	Shallow	Deep
Hydrogen (H)	1-12	13 or more
Carbon (C)	1-4	5 or more
Halogen (X)	1-2	3 or more
Sulfur (S)	1-2	3 or more
Nitrogen (N)	1-2	3 or more
Oxygen (O)	1-2	3 or more
Misc. (M)	1-2	3 or more

(4) *Ion Mass of Peaks.* All the rest of the holes in the card are devoted to recording the ion mass of the largest or most distinctive peaks in the mass spectrum. The parent mass peak is always included if it is 8 per cent or more of the base peak. The shallow punch positions code each mass value from 12 through 100. Deep punch positions carry the individual mass values up to 150. Thereafter, there is one punch position for each two mass values from 151-152 through 179-180, and for every ten masses from 181-190 through 371-380. Special holes code peaks between 381 and 400, 400 and 600, and 600 and 800. These values are all printed on the card to identify the proper holes.

Conventional Keysort sorting operations are used to arrange cards in any one of several possible orders, or to search for particular cards having specific sets of data for matching and identification purposes. A complete appreciation of what can be done with these, or any of the other punched card systems, can be had only after actually using the cards for some time.

The system developed at General Electric^{26, 27} is quite similar to the C.E.C. method just described. There is no provision for boiling points and although space is allotted for the coding of the elements in the compounds, it has not been used as yet. Punch positions are provided for indicating whether the data were obtained from a compound, a pyrolysis product or a mixture. A unique feature of the G. E. card is the affixing of the actual recorder tracing of a rough spectrum to the card.

Machine-Sorted (IBM) Cards. A system for indexing mass spectral and chemical structure data into IBM cards for sorting and correlating purposes was proposed by workers at Wyandotte Chemicals Corporation.²⁴ This card was identical to the previously described Wyandotte-A.S.T.M. card for indexing infrared absorption data, except that provision was made for indexing the mass spectrum peaks, the strongest peak and the molecular

²⁶ Zemany, P. D., "Punched Card Catalog of Mass Spectra Useful in Qualitative Analysis," *Analytical Chemistry*, **22**, 920-22 (July 1950).

²⁷ Zemany, P. D., "Identification of Complex Organic Materials," *Analytical Chemistry*, **24**, 1709-13 (November, 1952).

weight. The further development of the card and system was assumed by Subcommittee IV of A.S.T.M. Committee E-14 on Mass Spectrometry. There resulted a detailed proposal by workers at M. W. Kellogg Company²⁸ which increased the punching resolution, provided for both a base and parent peak, incorporated a more complete molecular formula and modified the chemical structure codes to meet the more limited classes of compounds susceptible to being handled in mass spectrographs. This was followed by a proposal by workers at Dow Chemical Company²⁹ which, although it makes use of IBM cards, is designed as a handsort or search file. The card carried the serial number, molecular weight, boiling point, number of chlorine and bromine atoms together with the mass numbers of the ten highest peaks, five other peculiar or particular peaks and as many as seven highest fractional peaks. Then, as many copies of each card are made as there are peaks punched into it. The cards are sorted and collated into blocks containing cards that have common mass numbers, then within each mass number block the cards are arranged according to the relative height of the peak on the particular card, and finally the blocks are arranged in order of the mass number. Thus, a copy of each card will be found in each mass number block for which it has a coded peak. With such a file one can go directly by hand to extract all cards indexing compounds whose highest peak has a given value or can also include all compounds that have an indexed peak at the given value regardless of its relative height. This achieves the results usually obtained by a first sorting operation at the expense of increasing the number of cards in the file by a factor of 10 or more. However, the comparison of cards in a given block to correlate the several peaks in a given spectrum with a given compound, when as many as 15 peaks may be involved, could become a rather complicated hand-sorting operation.

Since A.S.T.M. Committee E-14 has not taken official action on any system involving IBM cards and none are commercially available from other sources, a detailed description of the codes and procedures of such systems as are under consideration is not advisable for time may soon render any one obsolete.

Empirical Formula-Name Index

With the rapid accumulation of thousands of Wyandotte-A.S.T.M. cards indexing the absorption spectral data and chemical structure of as many different chemical compounds, the problem of maintaining an alphabetical index of the names of these compounds became rather complex. The need

²⁸ McCrea, J. M., "A Proposed Indexing System for Mass Spectra," submitted to A.S.T.M. Committee E-14, Subcommittee IV on May 26, 1953.

²⁹ McLafferty, F. W., and Gohlke, R. S., "A New Punched-Card Filing System for Mass Spectra," presented at the A.S.T.M. Committee E-14 meeting on Mass Spectrometry, New Orleans, La., May 28, 1954.

for such an index resulted from the frequent desire to locate the spectrogram of a given compound without having to resort to sorting the chemical structure data punched in the spectral data index cards. The complexity and duplicity in naming organic compounds made it desirable to establish a system that did not rely primarily on the name. Through the cooperative efforts of workers at Wyandotte Chemicals Corporation and Eastman Kodak Company, a system was developed which makes use of the empirical formula and a name punched into the same IBM card. Although IBM cards are used it is only for convenience in the initial preparation and subsequent duplications, since the cards are used as a hand file. The system, described in detail below, has been adopted by the American Society for Testing Materials and is being distributed by them, together with the other cards previously described. Thus, every spectral data card for a given compound has a formula-name card bearing the serial number of the spectrogram which serves to locate the spectrogram in the user's files.

Formula Name Cards. The formula-name cards indexing chemical compounds are designed to provide a ready means of obtaining all of the information about a given compound that has been coded into any of the other Wyandotte-A.S.T.M. cards. They may be arranged by machine into numerical order of the spectrum serial number, the numerical order of the empirical formulas, or into alphabetical order of the names and then used as a hand file for entry by any of these arrangements. The name, empirical formula and serial number are printed on the card to facilitate hand use. There is one card for each compound each time information concerning the compound is indexed into a different spectral data card. Thus, by entering the file for the compound benzene, one will find cards giving the serial number of the infrared absorption spectrograms, the ultraviolet absorption spectrogram and any other sets of data as have been incorporated into the system.

The formula-name card (see Figure 9-9) is divided into the following areas

[illegible]

Figure 9-9. Wyandotte-A.S.T.M. Name-Formula Card.

for incorporating data:

- (1) Elements—columns 1 through 8
- (2) Empirical Formula—columns 9 through 22
- (3) Miscellaneous code—column 25
- (4) Compound Name—columns 26 through 65
- (5) Reserved by A.S.T.M.—columns 23 and 24
- (6) Reserved for Private Use—columns 66 through 70
- (7) Reference of Serial Number—columns 71 through 80.

Each of the areas will be discussed in sufficient detail to permit one to make general use of the cards. Additional information may be obtained from the A.S.T.M.³⁰

(1) *Elements*. The identity of every element in the compound being coded is indicated by punches in this section. The same elements code used on the Wyandotte-A.S.T.M. x-ray diffraction cards and on the inorganic section of the chemical classification code for the infrared absorption data cards (see Table 9-5) is employed here in column 1 through 8. Thus, the numbers in brackets give the column numbers, and actinium would be punched at the "0" position in column 1, etc. This section can be used to segregate cards according to particular elements.

(2) *Empirical Formula*. Columns 9 through 22 are used to record the empirical formula of the compound being indexed. These numbers are punched directly into the appropriate columns and then interpreted or printed by machine along the top of the card. The chemical symbol for each element is also printed on the card for ease in hand searching. Table 9-14 relates the elements involved, the columns and the printing positions for interpreting the numbers in the proper place on the card.

It will be noted that only the more common elements are included in the table. Numbers of atoms greater than can be punched into the columns provided are recorded as the highest number that can be punched. Polymers and indeterminate structures receive no empirical formula punch. The empirical formulas punched here include all elements involved in the compound except water of hydration. Salts such as aniline hydrochloride, sometimes recorded as $C_6H_5NH_2 \cdot HCl$ would be punched into the card as C_6H_5ClN . An additional code in the next section of the card has been provided to indicate that such salts are involved.

(3) *Miscellaneous Code*. Column 25 provides for coding salts where the information would not be apparent from the combined empirical formula given on the card, it indicates when the compound coded is inorganic as well as the presences of water of hydration, which is not included in the

³⁰ "Codes and Instructions for A.S.T.M. Empirical Formula-Name Index Cards," *Ibid*, (1956).

TABLE 9-14

Column	Element	Printing Position
9-10	C	1-2
11-12	H	4-5
13	Br	7
14	Cl	9
15	F	11
16	I	13
17-18	N	15-16
19-20	O	18-19
21	S	21
22	Si	23

empirical formula. The miscellaneous code follows:

25-y HCl
 25-x HBr
 25-0 H₂SO₄
 25-1 Acetate
 25-2 Oxalate
 25-3 Phosphate
 25-4 Ammonium
 25-5 Nitrate
 25-6 H₂O (hydration)
 25-7 Inorganic
 25-8
 25-9 Other Acid Salt.

(4) *Compound Name.* The name of the compound, as closely as can be approached within the limitations of standard IBM equipment, is punched into columns 26 through 64. Since only capital letters, digits, comma, dash and slanting line are normally available, the names are printed in a modified but readily recognizable form in most cases. The "inversion" naming system²¹ as used by Chemical Abstracts is favored for use on these cards. However, no attempt has been made to rename all compounds by the Chemical Abstracts System. Such names as were supplied by authors were merely rearranged, applying the Chemical Abstracts principles, so that an "index name" could be used for alphabetizing. With these cards the empirical formula is all important and the name can be considered trivial. One need only be able to recognize any one of the possible names of the compound being searched for.

The first letter of the "index name" is always punched into column 30. Columns 26 through 29 provide for digits and/or letters which usually precede an index name but take no part in determining the alphabetical

²¹ "The Naming and Indexing of Chemical Compounds," *Chemical Abstracts*, 39, No. 24 (Introduction to the 1954 Subject Index).

sequence. If there are more such characters than can be accommodated in four columns, they are placed at the end of the name followed by a dash. Thus, 1,4,5,8-Naphthalenetetrol, 3-chloro is printed as NAPHTHALENETETROL, 3-CHLORO-1,4,5,8-. Greek letters are either spelled out or are represented by English equivalents for economy of space. The "prime mark" (') is indicated by the letters "PR". Thus, *o,o'*-Biphenol is printed as BIPHENOL, O,OPR-. Such other abbreviations as "C" for *cis*, "T" for *trans*, "M" for *meta*, "D" for dextrorotary, "N" for normal and many others which are perfectly obvious, are used in printing the names. Since parentheses are not available on the regular Model 552 interpreter, a slanting line has been used. Thus, a name written as 2-(methyldithio)-ethanol is printed on the card as ETHANOL, 2-/METHYLDITHIO/-. The same slanting lines must serve also as brackets. Every attempt is made to make the names as readable and correct as possible. The name is interpreted or printed in the lower printing space on the card in positions 5 through 44. On these cards an "x" overpunch produces the comma (,), a "y" overpunch produces the dash (-) and a combination of 0 and 1 punches in any column produces the slanting line (/).

If the name is too long to be punched into the 39 columns available on the first or parent card, it is broken at a normal position and the rest punched into the same columns of a second, or trailer, card which carries the same serial number. When this is done, a "T" is punched into column 65 of the parent card and a 9 into column 65 of the trailer. If the name cannot be punched into two cards, then a second trailer (or a third) may be used in which case both the letter T and the digit 9 are punched into column 65 of the middle trailers and only 9 into column 65 of the last trailer. All trailers carry the same serial number of the parent card but are different in color and carry no other punches or printing other than the portion of the name and the serial number. Punches into column 65 are interpreted to upper printing position 38.

In normal use, the cards are arranged strictly in numerical order of the number of atoms, and working from left (carbon atoms) to the right as printed across the top of the card. When no atoms of a particular kind are present, the fact is ignored and the next element to the right determines the order, but such cards are all placed behind the cards with formulas that *do* contain the element. Thus, the file begins with compounds containing one C and one H atom and all compounds containing one C and no H atoms fall behind those containing one C and the highest number of H atoms. This system is adhered to strictly. Otherwise, when atoms are present, the number of such atoms determines the position of the card and, working from the left, all cards having a given number of atoms are added to the file before cards containing a higher number of such atoms are included. All com-

pounds that do not contain carbon fall behind all compounds that do contain carbon and since this former group is chiefly inorganics it has been arranged in alphabetical order of the names. Polymers, trade name materials and all compounds that have no empirical formulas are filed alphabetically by name in the last section behind the inorganics.

A brief examination of the cards as they are distributed serves to familiarize one with the system. In this arrangement of the cards, it is convenient to file the trailer cards separately from the parent cards since they bear no empirical formula data, and to keep them in numerical order of the serial number so that they may readily be located when necessary.

(5) and (6). *Reserved Areas.* As on all other Wyandotte-A.S.T.M. cards, certain columns are reserved by A.S.T.M. for future use and another section is set aside for private use by individual laboratories. In the Formula-Name cards columns 23 and 24 are reserved by A.S.T.M. for their own purposes and columns 66 through 70 are available for private use.

(7) *Reference or Serial Number.* The serial numbers punched into these columns are the same as those in the corresponding infrared, ultraviolet or visible data cards for the compound. This includes the letters in columns 79 and 80 so that the designation on the card provides a direct reference to the location of the spectral data in the literature or local files.

Acknowledgment

The author wishes to thank Wyandotte Chemicals Corporation and the American Society for Testing Materials for permission to publish major sections of this work.

Chapter 10

AN APPLICATION OF RANDOM CODES FOR LITERATURE SEARCHING

CLAIRE K. SCHULTZ*

*Librarian, Merck Sharp & Dohme Research Laboratories
West Point, Pennsylvania*

Introduction

The random coding technique for indexing journal references has been employed in the Sharp & Dohme library since 1950. The library now indexes about 15,000 articles per year; a small amount with respect to the needs of some of this book's audience, but probably an "average" volume for special libraries attempting to index current literature for their organizations.

This library's literature service has to satisfy a group whose interests touch on nearly every phase of the biological, medical, and chemical sciences. The technique evolved for coding the names of diseases affecting man and domestic animals is quite specific, as is that for coding organic chemicals of known structure. Additional subject description for a given paper is supplied from a nonclassified, alphabetically arranged list of subject words.

The system can still be regarded as an experiment, in that ways for improving it are always under consideration. The virtues of a system that can be changed without invalidating any previous input must be recognized.

A staff of three typists and two professional people handle the input and output of the system.

The Conception of the System

The first consideration of punched cards for library use made it clear that a change from conventional indexing to a punched card system would be desirable only if the new system could provide all the functions of a standard index and also offer significant advantages. A system was needed that was capable of storing enough information to define and describe a reference (a standard index can do this) and that could also retrieve information quickly and easily by associating any and all subject fragments presented at the time of a search (a standard index cannot do this in many instances).

* *Present address:* Univac Division, Sperry Rand Corp., Philadelphia, Pa.

The punched card system adopted by the library had to be capable of handling a considerable volume of references, in terms of standard index systems, and also of meeting the diversified subject needs of the scientists it served. It was felt that in designing the library's application of punched cards an attempt should be made to get as much information as possible on one card, in order to make correlation of that information as easy and as meaningful as possible. It was recognized that the use of random superimposed symbols needed less card space than any other coding technique.

Continuing this reasoning, the point was reached where experiments could begin. Journal articles were coded from a dictionary of subject words arranged in alphabetical order. Each word had been assigned a random code number that represented four holes in a punched card, the codes for all of the subject words were superimposed in a field of ten columns on the card. An additional ten columns were used for coding the journal name, author and date of the reference. At that time IBM could not offer a machine to search for random superimposed codes, so a Remington Rand Sorter had to be used.

The enlarged and refined punched card system in operation at present, employing an IBM 101 Statistical Machine for searching (Figure 10-1), has grown from the base set by this 1950 experiment.

Random Codes

Mathematical discourses on coding systems, including random codes, are to be found in the literature^{1, 2}. A few lay observations stemming from the application of random codes will be presented here.

It has been pointed out that the technique of superimposition of random symbols offers the advantage that many "bits" of information can be coded into only a few columns of a card. Inherent in this technique is the possibility of creating false selections. There are numerous factors that modify this latter fact, some of which can be mentioned here.

The number of punches assigned to a code is one of the basic considerations. The more definitive the code, i.e., the more punches assigned to a code, the less probability there is of synthesizing it by chance when searching. However, the more punches used to define a term, the fewer the number of terms that can be superimposed into a field before it becomes saturated¹. This often has definite practical significance. Also the longer the code, the more cumbersome it is to work with in the clerical sense, and thus the greater the amount of human error that can be expected to enter the system.

¹ Calvin N. Mooers. "Putting probability to work in coding punched cards—Zatocoding." Presented before the Division of Chemical Education at the 112th meeting of the American Chemical Society, New York, Sept. 15-19, 1947.

² Carl S. Wise. Mathematical analysis of coding systems (Chapter 21, This Book).

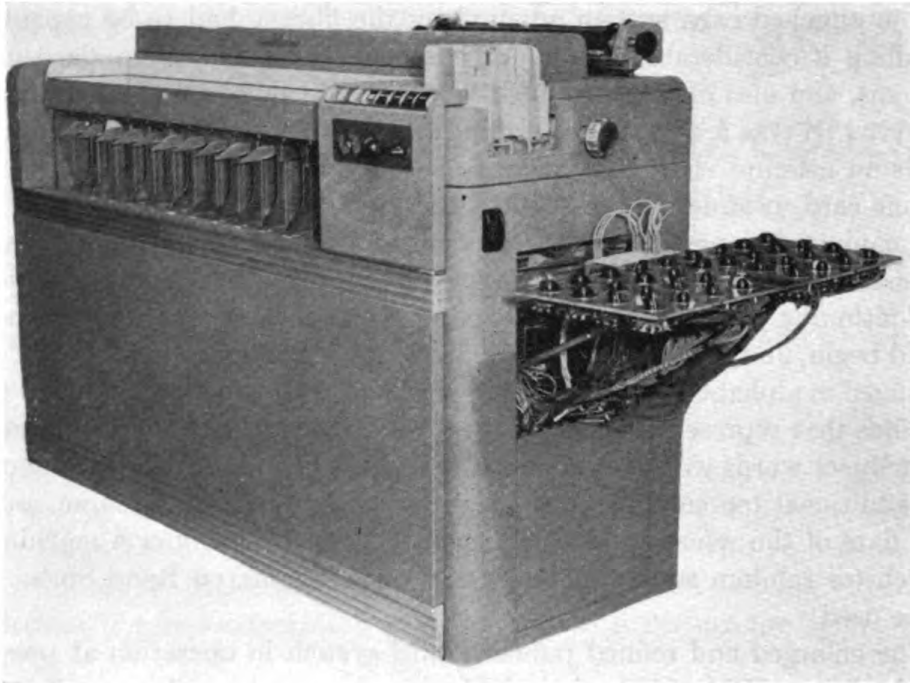


Figure 10-1. Electronic statistical machine Type 101, with auxiliary dial board.

Taking all of this into account, a code of four punches was decided upon for the application being described. Four-punch codes allow the superimposition of up to sixteen terms into a field of 100 punches. According to theory¹, not more than 69 per cent of the holes in any random field may be used up in coding information into that field. In this application a field contains 100 punching positions; so $69/4$ or 17 and a fraction terms, then, could be used per field. Actually, this number of terms becomes a little larger due to the overlapping of codes; for example, if the code 01-17-(49)-92 is punched and another term has the code 05-27-(49)-81, only three holes instead of four are needed to punch the second code into the card. However, to operate with a margin of safety, an upper limit of 16 terms per field has been observed here. In practice, the resulting false selection is small enough to cause insignificant interference. It is a natural practice to combine a group of terms for searching, rather than to look for a single one. This practice of amalgamating codes into a search "pattern" is an important contribution to the elimination of false selections.

Not all codes, though, are equally selective. The fact that they overlap, some by one digit, some by two digits, and some by three digits, makes them differ in their selectivity. Then, too, there is the fact that not all of the codes chosen for assignment to a dictionary will be used equally often. Some terms or groups of terms are needed more frequently than others for both indexing and searching. The codes for terms that are often used cut down

the selectivity of codes having numbers in common with them. Illustration:
In this library the terms:

humans
therapy
experimental
animals

occur frequently as a group on the punched card. The digits in their codes

08-10-36⁶⁷
97-86²¹-78
14⁴⁴-67-76
²³41-52-55

synthesize many other codes, one of which is traced in the above illustration. Searching for a term with this code would be almost useless because the volume of cards bearing these four codes would drop as false selections in prohibitive proportions. To alleviate that situation in this library, the terms used most frequently, either singly or in combination, were removed from the subject field and given direct punches in an otherwise unused portion of the card. Figure 10-2 shows these words and their punching positions. They can still be selected in combination with other codes used in the system but they no longer cut down the selectivity of other codes.

For the application being described, the list of random numbers assigned to the subject dictionary was derived from a table of random numbers in

										<table border="1"> <tr> <td>SUPP.</td> <td>TOPOGR.</td> <td>ETIOL.</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>			SUPP.	TOPOGR.	ETIOL.															
SUPP.	TOPOGR.	ETIOL.																												
☐ Human ☐ Experimental ☐ Vet. Medicine ☐ Other																														
1	2	3	4	5	6	7	8	9	10																					
☐ BLOOD	☐ PROT.	☐ ACT.	☐ ANT.	☐ VIT.	☐ HORM.	☐ CHILD																								
☐ HUMAN	☐ ANIM.	☐ EXP.	☐ VET.	☐ CHEM.	☐ BACT.	☐ VIR.	☐ THER.	☐ DIS.																						
31	32	33	34	35	36	37	38	39	40																					
☐ I. ARTC.	☐ I. ART.	☐ RES.	☐ I. S.	☐ I. P.	☐ S. C.	☐ ORAL	☐ AER.	☐ TOP.	☐ DOS.																					
☐ INJ.	☐ I. V.	☐ I. M.																												

Figure 10-2. Literature reference card before reference is entered.

Figure 10-3. Face of mark sensing card.

Figure 10-4. Back of mark sensing card.

Fisher and Yates³. That table lists the last 10 digits of a 20 place logarithmic table. Establishing a code for a field of ten columns (100 punching positions) allows the use of the numbers 00 through 99. (See the first 10 columns of Figure 10-3.) Constructing the code designating four punches in the field, therefore, required 8 digits. The first 8 digits of each entry in the Fisher and Yates table were utilized so that the entry 1324354657 yielded the code 13-24-35-46. The listings in the table were assigned to the alphabetically arranged list of subject headings, with no regard for establishing numerical relationships among the subject words.

Searching Random Codes by Machine

In the course of this library's program, experience has been accumulated with both the Remington Rand Sorter and the IBM Statistical Machine,

³ R. A. Fisher and P. Yates. Statistical tables for biological, agricultural and medical research. London, Oliver and Boyd, 1938.

Type 101. The operational aspects of these machines that are important to a system applying random codes can be summarized briefly.

The Remington Rand equipment comes with a sorting block that covers 144 punching positions (12 columns of the card). Various types of sorting pins make it possible to select a pattern of numbers within that area or to reject a pattern that might be associated with the pattern being selected. Selection into pockets is controlled by a bridge operating over only one column in any one pass through the machine. The decision as to what pocket will bear the product of the search, then, has to be a function of the pattern being searched. The Remington Rand card has 540 punching positions. The searching rate is 25,200 cards per hour. The machine will search for any pattern of punches that might be put into the 144 positions covered by the block. Aside from the electric motor driving the machine, it is completely mechanical in its sensing and selecting operations.

The IBM machine will search for a pattern of up to 60 holes anywhere on the card. Sequencing (either alphabetical or numerical) by means of the IBM 101 is achieved more easily and quickly than with the Remington Rand equipment; the machine also counts and prints. The sorting speed of the machine is 27,000 cards per hour. However, preference for the 101 has been based primarily on the increased amount of correlation possible at the time of the search.

To demonstrate this, one needs to consider the use of logical connectives in punched card sorting. A variety of logical patterns may describe the relationships of terms being used to formulate a search. For purposes of demonstration, the terms being searched can be designated as A, B, C, D, and can be assigned meaning as exemplified in the following:

- A. Benemid (therapeutic agent)
- B. Penicillin
- C. Gout
- D. Pneumonia

The logical connective *and* between each would signify that only those papers with reference to Benemid and penicillin being used in therapy of both gout and pneumonia are wanted ($A + B + C + D$). This type of searching, i.e., the use of the logical connective *and* can easily be accomplished by any punched card system.

If the request is for references to the therapy of pneumonia by these drugs, *but not* if the reference also concerns gout, the relationship is:

$$A + B + D - C$$

The use of the logical connective *but not* is accomplished in the Remington

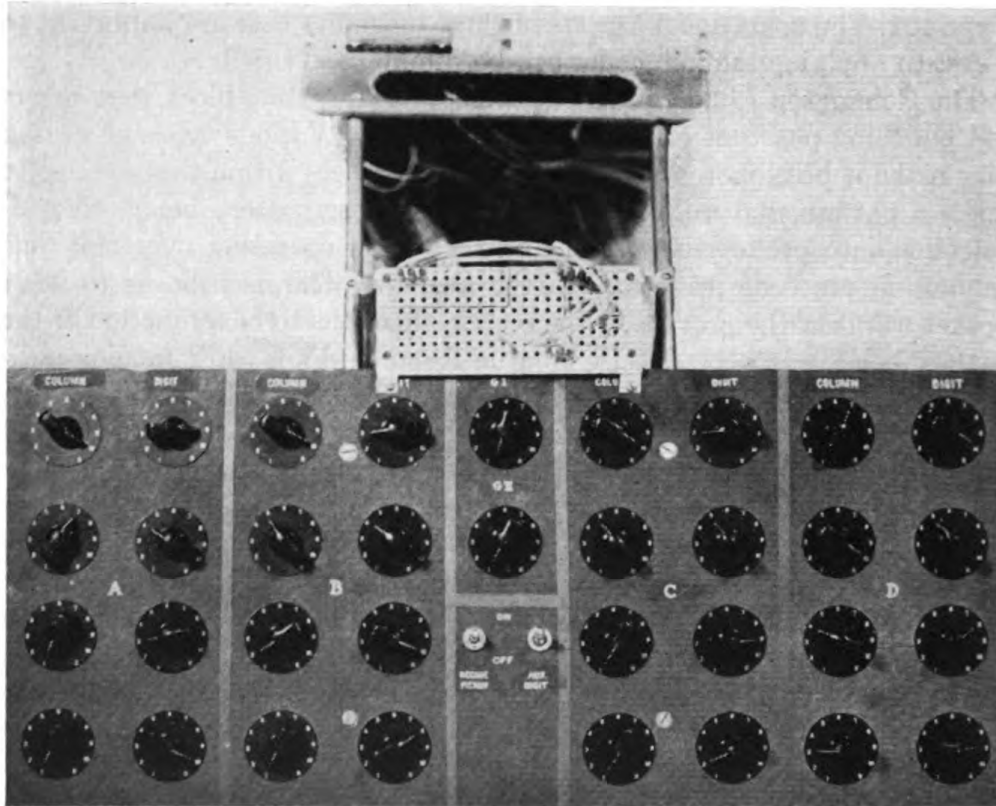


Figure 10-5. Close-up of an auxiliary dial board.

Rand sorter by using reject pins in the sorting block. With a hand-system, this type of search can be done only by selecting all four,

$$A + B + C + D$$

and then removing C from the pack selected in the first sort. The simple way in which the IBM accomplishes the search $A + B + D - C$ will become clear as the discussion progresses. Other logical connectives that can be dealt with only by an electronic sorter are exemplified by: *either-or*; *if-also*; and *if*.

The IBM equipment is so flexible that all of the combinations representing the logical connectives among the terms being searched are readily separable in routine operations. To facilitate these separations a wiring system has been developed to deliver the 15 possible combinations of A B C D each time a search is made⁴. The use of a dial board obviates the time and technical training needed to wire a control panel for a search,

⁴ The principle which led to this development of the auxiliary panel board was conceived by Mr. Bruse Moncrieff and his associates at the Home Office of the Prudential Life Insurance Company, Newark 1, New Jersey.

and through its use, it makes the 101 a more practical and efficient tool for literature searching. The dial board allows the codes for A, B, C, and D to be set in without requiring special knowledge or skill. Figure 10-5.

To continue with the example given above, the method of wiring used in this library delivers the answers to more questions than the specific one being asked; the answers to corollary questions are ready-made. An examination of what has dropped into each of the pockets of the machine as a result of this search will point up this fact.

Pocket	Combination	
12	ABCD	Papers making reference to Benemid, Penicillin, Gout and Pneumonia.
11	BCD	References to Penicillin, Gout and Pneumonia but not Benemid.
10	ACD	References to Benemid, Gout, Pneumonia, but not Penicillin.
1	ABC	References to Benemid, Penicillin, Gout, but not Pneumonia.
2	ABD	References to Benemid, Penicillin, Pneumonia, but not Gout.
3	CD	References to Gout and Pneumonia, but not to Penicillin or Benemid.
4	BD	References to Penicillin and Pneumonia, but not to Gout or Benemid.
5	BC	References to Penicillin and Gout, but not to Benemid or Pneumonia.
6	AC	References to Benemid and Gout, but not Penicillin or Pneumonia.
7	AD	References to Benemid and Pneumonia, but not Penicillin or Gout.
8	AB	References to Benemid and Penicillin, but not Gout or Pneumonia.
9	A	References to Benemid when it was not used in combination with Penicillin and when neither Gout nor Pneumonia were mentioned.
	B	References to Penicillin when neither A, C, nor D were present in the paper.
	C	References to Gout when neither A, B, nor D were present.
	D	References to Pneumonia when neither A, B, nor C were present.

The last four can be separated by passing the cards from pocket 9 through the machine again.

Dictionary

Important as they are, the three elements of a punched card system—the equipment for sorting, the design of the card, and the type of numbering system employed—can be looked on as tools for putting a subject dictionary into effect. Without a well constructed dictionary, the full values of punched cards for indexing literature cannot be realized.

There are two basic approaches to a dictionary with which to begin one's thinking: (1) an ordered classification in which the terms used and the

numbers assigned to them are correlated and dependent or (2) a nonclassified system where the dictionary is developed without regard to the logical relationships of terms. Example:

Classified Dictionary	Nonclassified Dictionary
animals	animals
mammals	dogs
dogs	mammals
rats	mumps
virus diseases	rats
mumps	virus diseases

For the classified dictionary, code numbers are usually assigned to emphasize logical relationships among entries, e.g.:

animals	2
mammals	2.2
dogs	2.21
rats	2.26
virus diseases	5.0
mumps	5.4

In a nonclassified random dictionary the entries are mutually independent and the assigned codes are of equal weight:

animals	14-35-27-48
dogs	16-19-88-92
mammals	01-08-36-99
mumps	02-08-16-31
rats	33-47-66-84
virus diseases	52-56-68-72

Terms may be incorporated into a non-classified dictionary as needed because there is no difficulty in making additions. There may be some difficulty in adding to a classified dictionary because its scope is more or less defined at the time it is set up. If one decided the next year to do research in three or four large fields not within the original scope of the system, the classified dictionary is likely to be under great stress.

One of the first things that becomes obvious in starting a non-classified dictionary is that it must never repeat any word or phrase. Classifications such as Dewey⁵, or Library of Congress⁶, use:

Religion—history
Medicine—history

⁵ Melvil Dewey. *Decimal classification and relative index*, 14th ed., New York, Lake Placid Club, 1942.

⁶ Martin, Nella Jane, ed., *Subject Headings used in the dictionary catalog of the Library of Congress*, 5th ed., Washington, Library of Congress, 1948.

Quarterly Cumulative Index Medicus List of Subject Headings⁷ uses:

Penicillin-toxicity
Sulfonamides-toxicity

The repetition of the word history or toxicity has no use in a nonclassified dictionary. Any word appearing in a nonclassified dictionary can be used in any combination desired; it appears there only once. To search for all the information in the file on *toxicity* would be as easy as searching for all the information in the file on *penicillin*. This is certainly not true of a standard card catalog where such a search would have to be carried through every drawer, toxicity being only a subdivision of the main headings.

Since repetition is unnecessary, the over-all size of the dictionary is considerably reduced and every term appearing in it has a utility not to be found in any other type of authority list. A word such as antagonism or anti can be used to form:

anti histamine
anti spasmodic
anti bacteria (antibacterial)
anti sepsis (antiseptic)
anti coagulation (anticoagulant)

In developing such a dictionary, rigorous attention must be given to the exclusion of synonyms and closely related terms. After completing a search, one does not want to discover that he should have asked the machine to select the cards bearing codes for *kittens* and *felines* as well as *cats*. These entries must be cross referenced and must never be assigned separate code numbers.

The present working dictionary in the system being described is a non-classified list of approximately 1000 indexing terms, consolidated as shown in the excerpt below:

Random number	Subject word	Remarks
12-15-29-91	ANTI HISTAMINES	
06-24-25-28	ANTIMONY AND ANTIMONY COMPOUNDS	
	ANTIPYRETICS	use: Fever; Therapy
	ANTISEPSIS	use: Antiseptics
36-46-54-82	ANTISEPTICS	also coded: Names of specific agents appearing in dictionary: Sterilization
	ANTISERUM	use: Anti; Serum; Immunity
	ANTITOXIN	use: Toxin; Anti; Immunity

⁷ American Medical Association. Quarterly Cumulative Index Medicus Subject Headings and Cross References, 2nd ed., Chicago, 1940.

This dictionary has thus far proved adequate for indexing some 30,000 references. As seen in the example given, not every subject word listed has a code number. In many cases the subject word is expressed by a group of other coded terms as shown opposite ANTISERUM. The term itself does not have a specific code but is found by searching for the cards containing the codes for anti, and serum, and immunity. At first glance, this may seem cumbersome, but the dividends are to be found in considering searches for terms other than antiserum when the reference is desired. That is, if all of the references pertaining to immunity were desired, the cards referring to antisera would be among those selected.

In some cases terms used by themselves have less selective power than is true of individual terms in a standard index. With the present system, a word such as CELLS may be used with any body organ or tissue. *BLOOD CELLS*, *PANCREATIC CELLS*, or *SERTOLI CELLS* all have the same code number for CELLS, but are distinguished by the additional codes for BLOOD, PANCREAS, and TESTIS. The reason for this, again, is to make it possible to find the reference under a greater diversity of searching conditions. The indexer has to think generically and specifically about every reference handled if this system is to approach the ideal in usefulness. For a paper about the stomach, the indexer would use not only *stomach* (specific), but also *gastrointestinal tract* (generic); the reason being that a researcher might be studying the effect of a certain drug on the gastrointestinal tract in general. With this type of coding he can get the papers on the subject without searching for esophagus, stomach, intestines, etc. It would be a mistake, though, to force the reader looking for references on the stomach to hand sort the entire pack of cards on gastrointestinal tract; references are conveniently indexed both ways, since to do so does not involve the preparation of more than one card.

All of the thinking about the generic and specific relationships of each term has to be set forth in the dictionary if the indexing is to be consistent. If the indexer uses just *penicillin* one time in indexing an article and then *antibiotics* and *penicillin* the next, the file cannot be expected to yield all the papers on antibiotics when searched. Anyone used to thinking in terms of the standard indexing practice may forget to index antibiotics on the penicillin paper; the dictionary will remind him to do so when he looks up the code for penicillin.

This same point could be made from many facets of the coding and searching. The indexer might analyze a paper very carefully and select the subjects indicating that the article gives information about the treatment of a disease in a child and that a certain dosage of a compound is given; but if the searcher wants all the papers on the use of that compound in human beings, he will not find the paper if it is coded only under children. It must

be coded under both *humans* and *children* and the dictionary must tell the coder:

Children
Humans

Also coded: humans.
Also coded: children, when pertinent.

Thinking in terms of requests from the information file, searching will be simplified if the dictionary anticipates similarities of terminology as much as possible. One person might ask for a search on diagnosis of diabetes, another for tests for blood sugar. These requests overlap somewhat, and if both words, diagnosis and tests, are coded by a different number, only a fraction of the wanted references will be obtained when searching for one of them. To take care of such words that are not synonyms, but which should not be given individual codes, our dictionary lists each of them in its alphabetical place, using the same code number for each of them.

In the development of the dictionary, the Quarterly Cumulative Index Medicus List of Subject Headings provided a frame of reference. Q.C.I.M. had been used as the library's authority list for two years previous to compiling the punched card dictionary. All the terms in Q.C.I.M. were considered for use with the punched cards, but the fact that the subject headings used by the library during the preceding two years had been checked made it easier to predict future needs. In addition to Q.C.I.M. coverage, many of the key research people reviewed an early version of the dictionary and suggested additional terminology which they felt necessary to cover their special fields of interest. The librarian correlated all of the suggestions and made the additions, cross references, and appropriate notes for the dictionary.

This random number dictionary has always worked well within the limits set for it. It was found by experience, however, that the volume of drug names and disease states encountered in the references indexed by this library was larger than had been anticipated when the dictionary was built. The system was suffering from not being specific enough in these two areas.

To correct the problem created by the drugs, it was decided to code them in a separate field from the other terms in the dictionary, and to assign every organic chemical of known structure a random number as it was encountered in the literature*. This was begun in 1953. In the three years that have ensued, the number of individual compounds coded is a little less than one thousand. The methodology for using the "chemical field" is the same as for the "subject field." In fact, the same group of random numbers was

* This statement needs to be modified to some extent. When it seems more expedient, salts are coded as the parent acid. In this sense, a random number can denote a group of compounds.

reused for assignment to compounds. Because they occur in separate fields, this is workable.

Diseases could have been taken out of the original dictionary and handled in the same way as compounds, or they could have been elaborated more thoroughly within the original dictionary. An intracompany development, though, made a third choice more expeditious. It was decided, for the sake of better intracompany communication, to adopt the American Medical Association's Standard Nomenclature of Diseases⁸, used by another of the Merck literature groups, for coding disease information. This is classified rather than a nonclassified list; the numbers are not random and cannot be superimposed to the same extent as random codes. Experience shows, though, that several diseases can be superimposed without serious problems of false selection. After six months of experience with this technique, it appears that it will save searching time for searches involving very specific diseases. Indexing time is lengthened because of the need to consult both SN⁸ and the original dictionary, and because of the intricacies of some of the decisions necessary for proper consistency in the use of SN. The 101 continues to search satisfactorily, in one pass through the machine, the combinations of searching terms needed from any or all of the three fields used for coding subjects, diseases, or drugs.

Cards and Card Design

Mark-sensing cards are used for preparation of the sorting index. Figures 10-3 and 10-4 show the two sides of the mark-sensing card as it is utilized. Perhaps it should be explained that the numerals of a mark-sensing card are enlarged and that one side of the card represents only 27 columns of a standard sorting card. Therefore, the punches resulting from the pencil marks placed on both the front and back of a mark-sensing card, 54 columns in all, fit easily into a standard punched card of 80 columns (Figure 10-6).

After being coded, the mark-sensing cards are run through an electrically operated punch that is activated by the graphite on the cards. The mark-sensing card is punched and used to prepare two duplicates of itself. It is then discarded; it cannot be used for searching purposes because the pencil marks interfere with the reading of its punches, making sorting inaccurate.

Two identically punched cards are prepared for each reference so that one can be used for sorting and the other can become part of the serially filed deck used for maintaining the system. The chief use of the latter is as a master deck for punching replacement cards when they are needed.

Figures 10-3 and 10-4 can be analyzed more closely to demonstrate how the coding is applied to the card. There are two random fields (columns

⁸ American Medical Association. Standard Nomenclature of Diseases and Operations. 4th ed. Philadelphia, Blakiston, 1952.

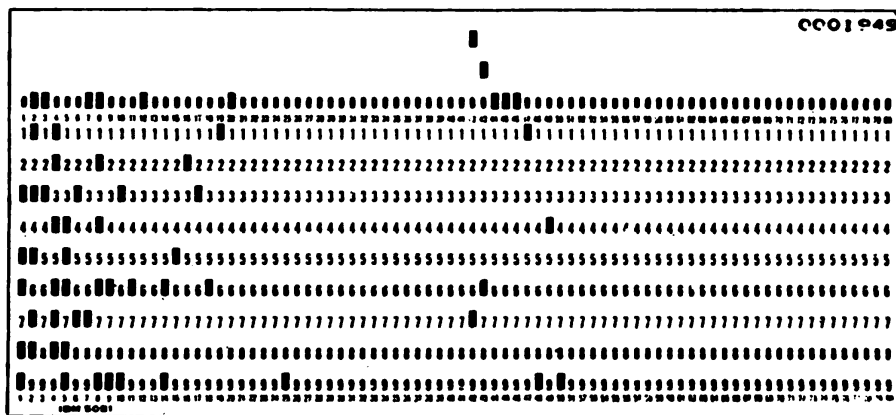


Figure 10-6. Standard IBM punched card with literature reference coded.

1-10, 31-40) of 10 columns each, devoted respectively to subject words and to the names of chemical compounds, as described previously. The names of authors are coded in columns 11-14. The first four consonants of the name form the code. The last two digits of the year of the reference are punched directly into columns 15-16. The journal, or other source for the reference, is coded in columns 17-20. Journal codes were devised by assigning four digit numbers in spaced sequence to an alphabetical list of the journals taken by the library. Example:

0160 *American Chemical Society, Journal.*
 0170 *American Dental Association, Journal.*
 0180 *American Documentation*
 0190 *American Drug Manufacturers' Association, Proceedings.*

Column 41 is labeled "special." It is used for direct punches to indicate a paper authored by a company staff member, a paper about a company product, and the color categories of the card that are to be used to record the reference. An explanation of the latter will follow. Columns 42-50 are used for coding the disease classification explained earlier in this chapter. The serial number of the reference (serial numbers assigned as references are indexed) is punched in columns 23-27.

Keeping the sorting deck in any kind of order is unnecessary until the volume of cards begins to make the searching time too long. The amount of time considered necessary varies from search to search, depending on the circumstances. However, in this library the goal is to keep machine time for an "ordinary" search to 15 minutes. It was recognized from the beginning that in dealing with journal references it would be useful to divide the sorting deck chronologically. This is done, in increments of one year, but even one year's accumulation (15,000 cards) is too much to sort every time a question is asked. To obviate this, colored cards have been used for

punching the sorting deck. Four colors are used to denote: (1) human clinical papers, (2) veterinary clinical papers, (3) experimental papers in the field of biology, (4) all others. The cards are filed in blocks, according to color, within each yearly division. The categories that the colors denote were chosen as being representative of the types into which most of the questions received can be divided. For some questions more than one of the color categories needs to be searched.

Getting The Reference into the System and Out Again

Persons desirous of knowing how the system operates may be interested in the following details:

Steps in Indexing and Coding a Journal Reference:

Professional Personnel

1. The indexer reads the title, the first paragraph and the summary of the article, carefully scanning the body of it.
2. a. Subject words (up to 16) are assigned from the subject dictionary to describe the article.
b. Diseases mentioned are looked up in SN.⁸ Their classification numbers are written in the box provided on the reference card shown in Figure 10-2.
c. The chemical compounds or trade names of compounds are noted for coding in the chemical field (there is an authority file kept on a Wheelindex).
d. The group of terms printed on the bottom of the reference card (Figure 10-2) are scanned. All terms that pertain to the article are indicated by a check mark.

Clerical Personnel

3. The reference card (Figure 10-2) is typed, giving the complete reference and the subject tracing for the article. This card bears a serial number, under which it is filed.
4. The code numbers are obtained for the terms indexed and a mark-sensing card is prepared by making pencil marks over each number to be punched. These marked cards are punched automatically by a mark-sensing punch.
5. The punched cards are checked for errors and then filed.

Steps in Making a Search:

Professional Personnel

1. The reference librarian translates the question into the terms of the system, choosing the most definitive ones for the search.

Professional or Clerical Personnel

2. The necessary code numbers are obtained.

3. These codes are set on the selection panel. The machine then delivers combinations of the terms being searched, as explained previously. The machine can also be used to sequence the selected cards by serial number and to print a list of those serial numbers.

Summary

This chapter describes an application of random superimposed coding for indexing journal references in the fields of medicine, pharmacology, and the allied sciences. The machine employed is the IBM Statistical Machine, Type 101.

Random coding is used for indexing chemical compounds and subject words describing a reference. In conjunction with this, a classification system is used for coding the names of diseases. All the information about an article including authors, date, and source is punched into one card. Any combination of the "bits" of information punched into a card can be examined in one pass through the machine. An auxiliary panel that makes it possible to set such search codes by means of a dial system is shown by photographs.

The indexing technique is described in detail in an attempt to illustrate some of the functional aspects of descriptive indexing. Searching is made more productive by the use of an auxiliary panel board which delivers the available variations of the information sought into separate pockets of the machine.

Chapter 11

SEARCHING METALLURGICAL LITERATURE

ALLEN KENT AND JAMES W. PERRY

*Center for Documentation and Communication Research
Western Reserve University, Cleveland, Ohio*

During 1957, a novel pilot searching service was initiated in the field of metallurgy. An experimental literature searching machine is being used to scan a file of encoded abstracts in response to questions submitted by various industrial and governmental organizations.

Development of this information service is based on certain processing methods and underlying principles which will be discussed under several headings as follows: history of the project; codes used; searching equipment; questions, their analysis and programming.

History of the Project and Introduction

The problems of coping with the increasing amount and complexity of scientific and technical literature which are facing users of metallurgical knowledge have long been a source of concern to the American Society for Metals.

The American Society for Metals recognized the need for bibliographic control more than a decade ago and took a first step by establishing the *ASM Review of Metal Literature* in 1944. At present, this is an abstracting service of the indicative rather than informative type which emphasizes the factors of promptness and completeness, without being exhaustive. A second step in the ASM program was the compilation and publication of the ASM-SLA Classification of Metallurgical Literature in 1950. Although the classification system by itself is a tool for organizing literature resources, it is specifically designed for use with a hand-sorted punched card system (see Chapter 5).

Both the *ASM Review of Metal Literature* and the ASM-SLA Classification were designed with the needs of the individual metallurgist particularly in view. Both services, however, immediately caught the attention of librarians and others who specialize in literature organization and searching, and the demand for something still more effective on a larger scale soon became evident to ASM. The hand-sorted punched card system is very well suited to collections up to about 10,000 documents. To handle the much larger collections that are encountered in metallurgical literature, a Committee on Mechanized Literature Searching appointed by the Board

of Trustees of the American Society for Metals, recommended that ASM sponsor a pilot operation to demonstrate the feasibility and advantages of applying computing-type equipment to the retrieval and correlation of metallurgical literature.

It was decided by the committee that the need for better methods of retrieving and correlating metallurgical literature was urgent, since the time was rapidly approaching when it would be cheaper to do a research job than to spend the time, effort and money required to do an adequate literature search.

The Center for Documentation and Communication Research, late in 1955, with a grant of \$75,000 from the American Society for Metals, undertook a five-year program to test and demonstrate the feasibility and usefulness of a mechanized searching service to ASM members. To achieve this purpose, a pilot-plant operation was required. The basis for such an operation had been provided during the past ten years both in equipment and also in new methods for indexing and coding information preparatory to machine searching.

Highlights of the pilot operation are as follows:

(1) Approximately 25,000 important metallurgical papers are being processed as the basis for pilot plant test and demonstration.

(2) Encoded "abstracts" are being used as the basis for searching and selecting operations. The "abstracts" used are telegraphic in character and they are particularly suitable for encoding for machine searching¹.

(3) The encoding of the abstracts for the 25,000 published papers is being conducted in such a way that a wide range of equipment can be used to conduct searching, selecting, and correlating operations.

(4) The editing to produce telegraphic-style abstracts and their subsequent encoding are based on techniques that make explicit for searching purposes both the generic significance and the specialized meaning of the terminology used in individual abstracts to express important aspects of

¹ See, for example, J. W. Perry, Allen Kent, and M. M. Berry, "Machine Literature Searching," pp 100-108, New York, Interscience, 1956; Allen Kent and J. W. Perry, "New Indexing-Abstracting System for Formal Reports, Development and Proof Services, Aberdeen Proving Ground," *Am. Doc.*, 8, No. 1, 34-36 (1957); J. W. Perry and Allen Kent, "The New Look in Library Science," *Appl. Mechanics Revs.*, 9, No. 11, 457-60 (1956); Allen Kent and C. R. Flagg, "Abstracting, Coding and Searching the Metallurgical Literature for ASM. The WRU Searching Selector," in J. H. Shera, A. Kent and J. W. Perry, eds., "Information Systems in Documentation," New York, Interscience, 1957; J. W. Perry and Allen Kent, "Tools for Machine Literature Searching: Semantic Code Dictionary; Applications; Searching Selector," New York, Interscience, (1958); M. R. Hyslop, "Inventory of Methods and Devices for Analysis, Storage and Retrieval of Information," in J. H. Shera, Allen Kent and J. W. Perry, eds., "Documentation in Action," pp. 128-130, New York, Reinhold, 1957.

subject matter. A code dictionary embracing about 10,000 frequently encountered scientific and technical terms is available and has been undergoing expansion to include metallurgical terminology.

(5) Informative abstracts from *Metallurgical Abstracts*, the *Journal of the Iron and Steel Institute*, and *Chemical Abstracts* are being used during the first two years of the program as the basis for preparing the encoded telegraphic-style abstracts for the pilot plant. Starting Sept. 1, 1957, original publications and papers for a test group are being used as the basis for the encoded abstracts (as well as for conventional abstracts).

(6) A limited searching service is being provided to ASM members at present (1958). (This will enable the market potential for the proposed service to be evaluated at a relatively early date. It is intended that this undertaking shall be placed on a self-supporting basis.)^{2a}

(7) The "pilot-plant" testing program is planned to extend over a total of five years. During the first two years attention was directed to the development phase.

Codes and Methods for Analysis

Encoding for machine searching requires that metallurgical information shall first be analyzed. An analyst reading an article can prepare both an abstract in a conventional form ready for publication and also a standardized telegraphic-style abstract ready to be encoded for machine searching¹. [Incidentally, at the same time that this is being performed, the analyst may also indicate what index entries are needed for the conventional subject index provided with the *ASM Review of Metal Literature*².]

Two aids have been provided the analyst who must perform this task³:

(1) A set of rules has been worked out for preparing standardized telegraphic abstracts in such a way as to eliminate the variations and complexities of English sentence structure.

(2) A series of subject matter analysis forms has been worked out to guide the consistent recording of important aspects of subject matter in the form of telegraphic abstracts. An example of completed analysis forms is given in Figure 11-1 (A-G). The italicized material given at the left-hand side of each part of the figure represents the headings presented on the

² See, for example, Allen Kent and J. W. Perry, "New Abstracting—Indexing System for Formal Reports, Development and Proof Services, Aberdeen Proving Ground," *Am. Doc.*, 7, 36–46 (1957); see also Chapter 6, in J. W. Perry and A. Kent, "Tools for Machine Literature Searching," Interscience, New York, 1958.

³ Jessica Melton, *Manual for Preparation of Telegraphic Abstracts*, Western Reserve University, Center for Documentation and Communication Research, Cleveland, March 25, 1957, (Multilithed); see also Chapter 5, in J. W. Perry and Allen Kent, "Tools for Machine Literature Searching," New York, Interscience, 1958.

^{2a} M. R. Hyslop, "Forecast of an Information Center," *Metal Progr.*, 74, 108–111 (July 1958).

<i>Form A</i>	
<i>Properties given for:</i>	Semiconductors,
<i>Material processed:</i>	Binary compounds,
<i>Starting material:</i>	Crystal/single/n-type
<i>Component:</i>	PbS, PbSe, PbTe
<i>Properties given:</i>	Semiconductivity
<i>Form B</i>	
<i>Material processed:</i>	Welds, Metal, Alloy
<i>Property influenced:</i>	Wear/mechanical, Abrasion
<i>Influenced by:</i>	Encounter # atoms, Area # contact
<i>Form C</i>	
<i>Process:</i>	Brazing/torch
<i>By means of:</i>	Flux
<i>Condition:</i>	Vapor
<i>Form D</i>	
<i>Material processed:</i>	Containers # bromine
<i>Component:</i>	Ni, Monel, Hastelloy, Pb, Steel, Teflon
<i>Testing technique:</i>	Immersion, Corrosion
<i>By means of:</i>	Bromine
<i>Condition:</i>	Wet, Dry
<i>Property determined:</i>	Resistivity # corrosion
<i>Form E</i>	
<i>Product:</i>	Alloy/N-155
<i>Component:</i>	Fe, Cr, Ni, Co
<i>Properties given:</i>	Physical
<i>Form F</i>	
<i>Machine or device:</i>	Vacuum furnace
<i>Rating, Size:</i>	Capacity 1,000 lbs.; Commercial
<i>Function:</i>	Melting
<i>Material processed:</i>	Steel
<i>Property influenced:</i>	Resistivity # temperature/high
<i>Form G</i>	
<i>Subassembly:</i>	X-ray unit/Seifert
<i>Rating, Size:</i>	Voltage/high
	Focus/fine
<i>Function:</i>	Measurement, Radiation
<i>Material processed:</i>	Metals
	Fe/gamma

Figure 11-1. Subject analysis forms for preparation of telegraphic abstracts. Italic headings at left-hand side of each part correspond to headings represented on the analysis forms; bold face material at right-hand side represents indexing information provided by analyst.

analysis forms; the material given in bold-face type at the right-hand side of each part of the figure, opposite the italicized headings, represents the index information provided by the analyst. It should be noted here that any combination or number of analysis forms may be used and that they may be altered as required to record adequately the subject matter of a

given document. Such alteration must remain, of course, within the provisions of the rules for generating the standardized telegraphic abstracts.

The next step in the process is to encode the individual terms and phrases of the telegraphic-style abstracts. A semantic code dictionary⁴ has been developed in which codes for specific terms express their meaning in such a way that related generic terms are made available as reference points for defining and conducting searching operations. For example, the code for "steel" will permit searches to be performed, not only to "steel," but also, more generically, to:

"metal alloys containing iron"
or to "alloys containing iron"
or to "ferrous metals"
or to "metals."

Also, the code for "length" will permit searches not only for "length," but also, more generically, to "material properties," or to "property."

The utility of this type of searching possibilities will become increasingly evident as the file of encoded abstracts continues to expand. Suffice it to say, for now, that sufficient flexibility and capacity are being provided for coping with diverse questions and with large files in a fashion not feasible with previous methods.

The code dictionary is maintained in the basic form of punched cards so that automatic procedures analogous to machine translation methods may be used to convert the telegraphic abstracts into encoded form.

The next step is to record the encoded abstract on punched paper tape—or other appropriate searching media, e.g., Minicards, magnetic tape, etc.

Searching Equipment

Several different machines can be used to accomplish searching of this type of material, as discussed earlier in the chapter. These include digital computers, computerlike devices such as the IBM X-794 and devices of the type exemplified by the WRU Searching Selector. These machines can be either specially programmed or are specifically designed to accomplish the type of searching to be described in the latter part of this chapter. In addition, the Eastman Kodak Minicard equipment can perform many types of searches that are made possible by encoding abstracts along the lines indicated above. Relatively minor electronic modifications would enable the Minicard equipment to perform the full range of searching operations that may be performed with encoded abstracts.

⁴J. W. Perry and Allen Kent, "Tools for Machine Literature Searching," New York, Interscience, 1958.

Digital computing equipment is, of course, available from the major business machine companies such as Sperry Rand (the series of machines known as Univac), International Business Machines (the "700" series), and others. The Minicard equipment referred to has been developed on Air Force funds by the Eastman Kodak Company in cooperation with the Magnavox Company of Indianapolis, Indiana. Announcement of the eventual availability and marketing of Minicard equipment was made as early as October 1955, in the *Wall Street Journal* and in other periodicals. The IBM X-794 was a commercial development that, it appears, could be made available within a reasonably short time.

The WRU Searching Selector was designed to search punched tape and specifically to perform the logical operations required for the effective searching of encoded abstracts. The WRU Searching Selector is characterized by very simple circuits, by unusual capabilities for performing up to ten simultaneous searches based on complex logical relationships, and by relatively low speed of operation.⁵

The WRU Searching Selector is able to perform the following functions⁶:

(1) Use patterns of holes in punched paper tape to record sequences of symbols. In this way, the characteristics of documents may be recorded one after another for subsequent search by the selector. (Individual symbols and combinations of symbols may be used to record the characteristics of documents in the same way that individual letters and combinations of letters are used to denote words in ordinary writing. It should also be noted that meaning may be ascribed to any combination of symbols as may be appropriate.)

(2) Read the tape by means of a Flexowriter and convert the patterns used to record successive symbols into corresponding electrical pulses, which then activate the discriminating unit.

(3) Detect those characteristics and combinations of characteristics which typify the subject contents of documents that are of pertinent interest. The discriminating unit is conditioned to detect such characteristics by appropriate wiring of the plug board prior to initiating a given search.

(4) Type out automatically the serial numbers of those documents whose characteristics correspond to the requirements of a given search. The scope of a search is expressed by specifying that the documents of pertinent interest shall have some one characteristic or some combination of char-

⁵ Design work is now under way for a high-speed counterpart of this searching selector. The scanning rate will be 200,000 to 300,000 abstracts per hour with 20 searches performed simultaneously.

⁶ Excerpted from J. W. Perry and Allen Kent, "The New Look in Library Science," *Appl. Mechanics Rev.*, 9, No. 11, 457-460 (November 1956). See also Chapter 18, in J. W. Perry and Allen Kent "Tools for Machine Literature Searching," New York, Interscience, 1958.

acteristics. Possibilities for specifying combinations of characteristics are outlined below.

The sequences of symbols may be organized into combinations analogous to "syllables" from which "words" may then be built up and from which, in turn, combinations analogous to "phrases," "sentences," and "paragraphs" may be built up. If the capital letters, A, B, C, D, etc., are used to designate individual symbols, then letters with subscripts may be used to designate various levels of combinations as follows:

A_1, B_1, C_1, D_1 , etc., for "syllables"
 A_2, B_2, C_2, D_2 , etc., for "words"
 A_3, B_3, C_3, D_3 , etc., for "phrases"
 A_4, B_4, C_4, D_4 , etc., for "sentences"
 A_5, B_5, C_5, D_5 , etc., for "paragraphs"
 A_6, B_6, C_6, D_6 , etc., for "messages".

This ability to organize characteristics into sets analogous to "phrases," "sentences," etc., is important in preventing false association of characteristics when searching. For example, by proper "phrasing" it is possible to prevent the properties of one alloy being incorrectly attributed to some other alloy.

At any level, which we may term the "n-th" level, each combination consists in general of a number of component combinations at the "n - 1" level. Each of several "n-th" level combinations, denoted by A_n, B_n, C_n, D_n , etc., may be specified in terms of component units designated by $A_{n-1}, B_{n-1}, C_{n-1}$, etc. Thus, in conducting a search, it may be specified as a condition that a document will be identified as being of pertinent interest, that at least one "n-th" level combination shall be characterized by certain component units. Specification of the component units may be set up on the basis of the following relationships. It may be specified that:

(1) *All of several components units must be present.* This requirement constitutes a logical product that may be symbolized, for example, by:

$$A_{n-1} \cdot B_{n-1} \cdot C_{n-1}, \text{ etc.}$$

In specifying logical products, further requirements as to order may be imposed. Thus, for example, it may be required that all components specified by a logical product shall occur in sequence. For example, it may be required that A_{n-1} shall be followed by B_{n-1} and it, in turn, by C_{n-1} . This requirement may be symbolized by

$$\langle A_{n-1} \cdot B_{n-1} \cdot C_{n-1} \rangle$$

The reverse order of these three components might also be specified as

denoted by

$$\langle C_{n-1} \cdot B_{n-1} \cdot A_{n-1} \rangle$$

(2) *Any one of several component units or, alternately, one or more of several component units must be present.* This requirement constitutes a logical sum that may be symbolized, for example, by

$$A_{n-1} + B_{n-1} + C_{n-1}, \text{ etc.}$$

(3) *At least one component unit must be present but at least one other component unit must be absent.* This requirement constitutes a logical difference that may be symbolized, for example, by

$$A_{n-1} - B_{n-1}$$

Here, also, order may be designated. Thus it may be specified that B_{n-1} may not follow A_{n-1} . This requirement would be symbolized by

$$\langle A_{n-1} - B_{n-1} \rangle$$

Alternately, it might be specified that B_{n-1} may not precede A_{n-1} and this would be symbolized by

$$\langle - B_{n-1} \cdot A_{n-1} \rangle$$

(4) *Combinations of component units expressed by complex logical relationships must be present.* Such logical relationships as the following may be specified

$$(A_{n-1} \cdot B_{n-1} - C_{n-1})(D_{n-1} + E_{n-1})$$

$$(A_{n-1} + B_{n-1})(C_{n-1} - D_{n-1}) E_{n-1}$$

$$(A_{n-1} - B_{n-1})(C_{n-1} \cdot D_{n-1} - E_{n-1})$$

Any such complex logical relationship may be set up as required at any level. Such complex logical relationships may also involve specification of sequential order. Using the symbols $\langle \rangle$ to denote order as before, we might specify such search requirements as

$$\langle (A_{n-1} \cdot B_{n-1} - C_{n-1}) \rangle (D_{n-1} + E_{n-1})$$

$$(A_{n-1} + B_{n-1}) \langle \langle (C_{n-1} - D_{n-1}) \rangle E_{n-1} \rangle$$

$$\langle \langle (A_{n-1} - B_{n-1}) \rangle (C_{n-1} \cdot D_{n-1} - E_{n-1}) \rangle$$

Application of these capabilities means that "syllables" may be specified in terms of component symbols, e.g., letters, "words" may be specified in terms of "syllables," "phrases" in terms of "words," "sentences" in terms of "phrases," "paragraphs" in terms of "sentences," and "messages" in terms of "paragraphs."

As stated above, the abstract formulation of higher order characteristics in terms of their lower order components has been restricted to the special case that the "n-th" order characteristic, e.g., a "sentence" shall be specified in terms of its components at the next lower "n-1" level, e.g., at the "phrase" level. The WRU Searching Selector may be readily programmed so that a higher order combination is specified in terms of any desired combination of lower order characteristics provided only that their order is less than "n". Thus, for example, "sentences" may be specified not only in terms of logically defined combinations whose component elements may be "phrases" and also "words", "syllables" and individual symbols. In abstract formulation, a characteristic of "n-th" order may be specified:

- (1) as a logical product e.g., $A_j \cdot B_k \cdot C_l$
- (2) as a logical sum, e.g., $A_j + B_k + C_l$
- (3) as a logical difference, e.g., $A_j - B_k$
- (4) as a complex logical combination e.g.,

$[(A_k \cdot B_g) - C_h] [D_f + E_l] (A_k \cdot B_g) (C_f - D_g) E_h [A_g - B_f] [(C_h \cdot D_k) - E_l]$
 where "e, f, g, h, j, k, l" are each less than "n".

In such combinations, two or more component elements of lower order than "n" may be of the same lesser order.

These capabilities enable encoded telegraphic abstracts to be searched conveniently and effectively. Various other machines with somewhat lesser capabilities for performing the searching operations required are available commercially from one or another business machine manufacturer, and are too numerous to enumerate here. For example, the precursor of the IBM 101 statistical machine—the Census 100—has some of the features required for searching encoded abstracts as prepared for the American Society for Metals. It must be noted, however, that with such machines modification of the encoded abstracts would probably be advisable to match the limited capabilities for literature searching by these machines.

Questions, Their Analysis and Programming

The question chosen as an example for detailed consideration is the following:

"How does the presence of vanadium in titanium alloys affect their casting?" This simple question permits some of the most important capabilities of searching system to be illustrated. In particular, the searching strategy may be varied depending on whether it is desired to extend the range of selected papers to include those that contain information that may be of less direct interest.

It is perhaps obvious that reports of experiments and tests directed to the casting of vanadium containing titanium alloys or to studies on their castability properties will be of prime pertinency to our example question. Accordingly, a sharply focused interpretation of our question might be formulated, with designation of code elements, as follows:

"Select those encoded abstracts that mention an alloy (LALL.001) whose principal component (KUJ) is titanium (MATL.11.□TQI) with a lesser component (KUJ), vanadium (MATL.11.□V), when the alloy is either the material processed (KEJ) by the process (KAM) casting (CUNS. 025) or related terms (CUNS.25X) or when a property given for (KOV) the alloy is castability (CUNS.25X.PAPR.004) which is designated either as a property given (KWV) or as a property influenced (KAP)."

Conversion of this statement of the example question into a searching machine program requires specification both of logical and also of sequential relationships between the above cited code elements, such as LALL.001, KUJ, etc. Such relationships may be expressed symbolically as follows:

Code elements (Syllable level)

$A_1 = \text{LALL.}$	$G_1 = \square V.$	$M_1 = \text{KOV.}$
$B_1 = \text{001.}$	$H_1 = \text{KEJ.}$	$N_1 = \text{PAPR.}$
$C_1 = \text{KUJ.}$	$I_1 = \text{KAM.}$	$O_1 = \text{004.}$
$D_1 = \text{MATL.}$	$J_1 = \text{CUNS.}$	$P_1 = \text{KWV.}$
$E_1 = \text{11.}$	$K_1 = \text{025.}$	$Q_1 = \text{KAP.}$
$F_1 = \square \text{TQI.}$	$L_1 = \text{25X.}$	

(Here the three letter codes with K as initial letter are role indicators which indicate mode of involvement of the term whose code follows within a subphrase.)

Subphrase level

$$\begin{aligned} A_2 &= \langle H_1 \cdot A_1 \cdot B_1 \rangle & C_2 &= \langle C_1 \cdot D_1 \cdot E_1 \cdot F_1 \rangle \\ B_2 &= \langle M_1 \cdot A_1 \cdot B_1 \rangle & D_2 &= \langle C_1 \cdot D_1 \cdot E_1 \cdot G_1 \rangle \\ E_2 &= \langle I_1 \cdot J_1 (K_1 + L_1) \rangle \\ F_2 &= \langle (P_1 + Q_1) \langle J_1 \cdot L_1 \cdot N_1 \cdot O_1 \rangle \rangle \end{aligned}$$

Phrase level

$$\begin{aligned} A_3 &= \langle A_2 \cdot C_2 \cdot D_2 \rangle & C_3 &= E_2 \\ B_3 &= \langle B_2 \cdot C_2 \cdot D_2 \rangle & D_3 &= F_2 \end{aligned}$$

Sentence level

$$A_4 = (A_3 \cdot C_3) + (B_3 \cdot D_3)$$

In this encoding of the search, particular importance attaches to the combination of code elements denoted by $E_2 = \langle I_1 \cdot J_1 (K_1 + L_1) \rangle$ that is

to the combination of code elements KAM, CUNS and either 025 or 25X when detected within a subphrase in that order. Here the role indicator KAM indicates that the term whose code follows in the same subphrase denotes a process. The combination of the code element CUNS with either 025 or 25X occurs in the codes for casting and various related terms denoting various specific casting processes, materials used in casting and related terms as may be evident by the following examples from the code dictionary,

continuous casting	CUNS.25X.CYNT.10X.MWTL.PASS.001
Junghaus-Rossi process	CUNS.25X.CYNT.10X.MWTL.PASS.002.
Asarco process	CUNS.25X.CYNT.10X.MWTL.PASS.005.
core oil	CUNS.25X.FATT.3X.MWPR.24X.MWTL.001.
foundry	CUNS.25X.LACN.001.
shelf molding	CUNS.25X.MWTL.PASS.003.
founding	CUNS.25X.MWTL.PASS.009.

as well as numerous additional terms for casting processes.

Thus specification of the combination of code elements CUNS. and either 025 or 25X is equivalent to citing a lengthy list of single terms relating to castings, while the requirement that KAM shall be found with CUNS and either 025 or 25X, in that order, effectively selects out that those terms that designate casting operations.

The combinations of code elements designated by C_2 and D_2 , namely MATL. and 11., indicates a class of metals in the ASM-SLA classification while $\square$ TQI and $\square$ V are special codes for the chemical elements titanium and vanadium. Note that specification of titanium as the principal component of the alloy is made possible by encoding alloys that the main metallic component is cited first among the components of an alloy.

The above presented formulation of our example question illustrates its conversion to a sharply focused machine selection program to identify those encoded abstracts that are virtually certain to be of direct pertinent interest. The scope of search may be readily extended, by various alterations in the machine searching program, to accomplish identification of additional encoded abstracts that may be expected to refer to information of less directly relevant interest. For example, the search might be extended to include:

"As products, castings, and the like comprising titanium alloys containing vanadium." (Here the combination of code elements, KWJ and C-NS, with either 025 or 25X within a single subphrase will characterize castings and similar products when mentioned as products.)

If R_1 is used to designate KWJ, the role indicator for "Product" and S_1

is used to designate C-NS, then this extension of the search may be symbolized as follows:

Subphrase level

$$G_2 = \langle R_1 \cdot S_1 (K_1 + L_1) \rangle$$

$$H_2 = \langle C_1 \cdot A_1 \cdot B_1 \rangle$$

Phrase level

$$E_3 = \langle G_2 \cdot H_2 \cdot C_2 \cdot D_2 \rangle \quad F_3 = G_2$$

The over-all scope of the extended search (denoted by B_4) may then be abstractly specified as follows:

Sentence level

$$B_4 = A_3 \cdot C_3 + B_3 \cdot D_3 + E_3 + A_3 \cdot F_3$$

A further extension of the scope of search may be made to include encoded abstracts which cite the casting or castability or castings (as products) of alloys mentioned by trade name only in the original publication. If such alloys are known to contain titanium as principal component and vanadium as lesser component, this will be indicated by their codes and the latter, in turn, enable the searching selector to be programmed to detect them by specifying a suitably defined sequence of code elements, namely; LALL and $\square TQI \square V$. In this way, the scope of search may be extended, if desired, to include papers which referred to the casting or castability of alloys whose trade names provided the only indication that they were titanium alloys containing vanadium. Thus, the previously extended search, denoted by B_4 may be further extended by setting up additional search requirements at various levels as follows:

Code elements (Syllable level)

$$T_1 = \square TQI \square V$$

Subphrase level

$$I_2 = \langle H_1 \cdot A_1 \cdot T_1 \rangle \quad J_2 = \langle M_1 \cdot A_1 \cdot T_1 \rangle$$

$$K_2 = \langle C_1 \cdot A_1 \cdot T_1 \rangle$$

Phrase level

$$G_3 = I_2 \quad H_3 = J_2 \quad I_3 = \langle G_2 \cdot K_2 \rangle$$

The overall scope of this further extended search (denoted by C_4) may then be abstractly specified as follows:

Sentence level

$$C_4 = [(C_3 + F_3) (A_3 + G_3)] + [D_3 (B_3 + H_3)] + E_3 + I_3$$

It should be understood that both the narrow and the broad interpretations of our example search, as represented by the logical formulations, A_4 , B_4 and C_4 above, may be searched simultaneously. The WRU Searching Selector is designed to search ten independent questions at once.

Result of Search

The basic operation of the WRU Searching Selector is the scanning of a continuous tape in which the punching of successive combinations of holes records the succession of symbols that spell out encoded abstracts. The scanning operation causes the search requirements to be matched with the encoded characteristics of the various abstracts.

As mentioned earlier, ten searches may be performed simultaneously. The machine automatically types out: (1) the serial number of each selected article; (2) the number of the search that has been satisfied; and (3) a bibliographic citation. For example, the machine may type out:

7325 1 G. H. Schippereit, R. M. Lang and J. C. Kura. American Foundrymen's Society, Transactions, 65, 499-512 (May, 1957).

to indicate that article 7325 in the file satisfies search number 1 of ten searches and that this article is located in *American Foundrymen's Society, Transactions*, Volume 65, May, 1957, pages 499-512.

After identification, articles are removed clerically from the file and presented to the expert analyst for review and evaluation. If the Minicard Selector is used instead, the cards selected can contain a microphoto of the original paper which may be viewed in a suitable reader.

Chapter 12

CLASSIFICATION, SEARCHING AND MECHANIZATION IN THE U.S. PATENT OFFICE

B. E. LANHAM AND J. LEIBOWITZ

U. S. Patent Office, Washington, D. C.

Introduction

This chapter includes references and descriptions of Patent Office features such as history, conventional patent searching and classification, as well as progress and experiments in mechanized searching. Its purpose is to provide the reader with a general overall view of the Patent Office and its functions and objectives. For those who desire more specific information the listed references may be of assistance.

Considerable interest has been manifested in the various operations of the Patent Office and its experimentation in patent search mechanization, and it is hoped that the included descriptive matter will be of value.

Historical Background of Patent Searching and Classification

On April 10, 1790, President Washington signed the Congressional Bill under provision of Article 1, Section 8, of the Constitution, authorizing the grant of patents by the U. S. Government.

The 1790 Act required as a condition precedent to the grant of a patent that satisfactory evidence of novelty, utility, and invention be established, which requirements are in existence at the present time. A "prior art search" was thus necessary, and since it was apparently limited to the relatively few patents issued by American Colonies and States as well as among books on mechanics and industrial arts, no need for classification of the searchable material was then necessary.

The first U. S. patent was issued on July 31, 1790, and the total was 57 on February 21, 1793, when a new Patent Act replaced the earlier one. The new Act substituted a "registration" system for the "examination" system, and that unfortunate replacement continued until the Act of July 4, 1836, was passed. At that date 9,957 patents had been issued and the new Act reestablished the examination system, including prior art searching—for that which had been invented or used before. A patent classification was formed, including 22 classes with no subclasses, the search material being

the manuscripts filed by applicants. The Patent Office was established as a distinct bureau with the appointment of the first Commissioner of Patents. Patent No. 1 of the present series started with the 1836 Act and was issued on July 13, 1836, and in 1866 the printing of patent copies was permanently started. At the end of 1868 slightly more than 80,000 patents had been issued; these were divided into 36 classes in alphabetical order of their titles, with some classes containing sections not in accord with present subclasses.

In 1872 the previous alphabetical classification was revised and the issued 131,000 patents distributed among its 145 classes. In 1880 publication of the first classification with both classes and subclasses occurred, with minor subclasses indented under major ones. None of these early classification systems, however, were based on the principles governing the allowance of patents, but they should have been devised on this basis.

The Classification Division (now Classification Group, consisting of five Divisions and a Service Branch) was established in 1898 by authorization of Congress^{1, 2}.

Patentability Requirements and Uses of Classification

In performance of its function the Patent Office examines applications to determine whether or not the applicants are entitled to patents under the law. In the language of the Statute, "Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of . . ." patentability as expressed in other sections of the Statute³. The purpose of the search is to determine whether the subject matter for which a patent is sought satisfies the statutory requirements for novelty and invention.

No quantitative yardstick for invention has been devised. Certain criteria have been developed, however, as a result of years of experience and in view of various appellate and judicial decisions which aid in determination of the question of invention. As a broad statement as to these factors, if the subject matter sought to be patented involves an obvious dissimilarity to the most nearly similar thing known, it is not regarded as inventive, and the same applies to advances which would be considered obvious to a person having ordinary skill in the art. The Examiner is thus interested not only in identical but in all related and analogous subject matter.

¹ "The Story of the United States Patent Office, 1790-1956," third edition. Superintendent of Documents, 25 cents.

² M. F. Bailey, "History of Classification of Patents," *J.P.O.S.*, 18, 463-507, 537-575 (July and August, 1946). (Reprints are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

³ "Patent Laws." Superintendent of Documents, 25 cents.

In addition to the searches made by the Examiner there are related types made by others. Before filing an application, an inventor usually makes a "pre-ex" search to determine the probable novelty of his invention. Those who question the validity of a patent and wish to prove in court that it was erroneously issued, make an exhaustive search in an attempt to find anticipatory references against the claims. Others are interested in a general study of patents to determine developments in certain fields of endeavor, known as "state of the art" search. The claims of unexpired patents are studied to determine if manufacture of a certain product or performance of a certain process may or may not result in an infringement suit. Research scientists find that patent literature may furnish valuable background material for a project and thus by finding what had already been achieved, avoid duplication of effort.

The Classification System

The Patent Office classification system is intended to provide facilities for storage and location of patents which relate to all branches of science and technology whereby searchers may, within a reasonable time, have available the art segments of interest. This purpose has been appreciably but not entirely accomplished.

At present there are approximately 309 classes which contain a total of over 52,000 subclasses. The number of U. S. patents issued is over 2,800,000 (this figure does not include reissue, design or plant patents). The class and subclass schedules are contained in the Manual of Classification⁴ which includes an alphabetical index of titles with reference to pertinent classes and subclasses. Revision of all the classes has been practically completed under modern methods, and all revised ones, including their subclasses, have definitions and notes as to content and scope of their subject matter and relationships, differences, and search suggestions as to other pertinent ones⁵.

The methods of classifying patents are quite complex and are sometimes considered inconsistent by inexperienced classifiers and searchers. The following examples illustrate a few general types.

All organic and inorganic chemical compounds, regardless of their disclosed utility, are classified on the basis of their chemical constitution. Most compositions of matter, i.e., mixtures of two or more ingredients, are classified primarily on the basis of their necessary functions or inherent properties rather than upon the basis of ingredients. Such primary groups

⁴ "Manual of Classification," plus Alphabetical Index. Superintendent of Documents, \$12.00.

⁵ "Definition Bulletins." Purchasable from Patent Office. Identification number and price of Bulletin obtainable upon receipt of class number. Ex.: Class 260, Bulletin No. 200, 80 cents.

are subdivided into subclasses on the secondary basis of selected ingredients. Processes of making compounds or compositions are classified, in most instances, on the basis of the resulting products. Processes, such as manufacturing (Class 29, Metal Working), the application of coating material to a base (Class 117, Coating, Processes and Miscellaneous Products), are classified on the basis of their function or ultimate result, and the characteristics of the subclasses are the selected operations included in the processes. Processes which do not result in a product (Class 209, Classifying, Separating and Assorting Solids) are classified on the same basis. In some classes, such as 209, the processes and apparatus for performing the same functions are classified together since the common search is consistent and coextensive.

Manufactured products are usually classified according to their disclosed and necessary function or utility (Class 2, Apparel; Class 81, Tools).

Machines are generally classified on the necessary mode of operation and effect produced rather than upon the specific material handled (Class 202, Distillation; Class 241, Solid Material Comminution or Disintegration). However a few types of machines are classified on the basis of the material handled (Class 80, Metal Rolling).

Class 241 is an illustration of a modern machine class. It contains three main sections. The first includes subject matter (the class machine) combined with features such as means to prevent explosions therein which are not necessary for the essential functioning of the machine. The second includes the machine per se, and the third subcombinations of the machine, such as disc grinding elements, which are not classified elsewhere. Other parts of Class 241 machines, such as motors, alloys, etc., not included in the class, involve other search fields.

As a general statement, a patent is classified on the basis of the inventive or claimed subject matter, and since consideration of patentability of claimed subject matter in an application is not limited to what has previously been patented, the major amount of unclaimed disclosures in a patent, as well as in other types of literature, are of searchable value. Thus cross-references of patents based on such disclosures are placed in pertinent classes and their subclasses.

More specific details of the foregoing references to general illustrative classification have been previously published^{2, 6-9}.

² "The Classification of Patents" (2d Revision). (Copies or reprints available from Research and Development, U. S. Patent Office, Washington 25, D. C. Copies of the first edition only are now available.)

⁷ M. C. Rosa, "Problems of Classifying Chemical Patents," *J.P.O.S.*, **19**, 241-261 (April, 1947).

⁸ B. E. Lanham, "Chemical Patent Searches," *Ind. Eng. Chem.*, **43**, 2494-2496, November, 1951; *J.P.O.S.*, **34**, 315-323 (May, 1952).

⁹ Bailey, M. F., B. E. Lanham and J. Leibowitz, "Problems of Classification and

The classification is based upon the criteria of patentability, and the following excerpt from the Manual of Classification illustrates its principles:

"As all patentable arts or instruments are created for an ulterior utility, the characteristic selected as the basis of classification is that of essential function or effect. Arts or instruments having like functions, producing like products, or achieving like effects are brought together; but the functions or effects that serve as a basis of classification must be proximate or essential, not remote or accidental."

The foregoing class examples illustrate what appear to be inconsistencies in establishment of classes. However the arrangement of various classes as related to different subjects is made on the basis of the ultimate property or utility to be searched as specified in the above quotation. Thus the inconsistencies are more apparent than real.

Conventional Patent Searching

The rather broad patent searching operations described here are intended as examples and suggestions to benefit those whose types of searches have been mentioned^{7, 8, 10-12}.

The problem in patent searching is not particularly different from others where it is desirable to isolate certain information from a vast and heterogeneous field of subject matter. As to questions of patentability, specific differences exist as to the types of relationships sought, the comprehensiveness of the subject matter and the variability in search requirements.

Prior to the start of a search, whether it is to be manual or mechanized, a thorough study of the subject matter of interest, as well as the purpose of the search, should be made. All aspects should be analyzed and verified and determination made as to whether the search is to be limited to the precise product, process or apparatus or to equivalents thereof, or to generic or specific variations or fragments thereof.

If the search is promoted by a desire to file an application for a patent, the Examiner's search viewpoints and patentability requirements should be reviewed. Fragments of disclosures of the claimed invention may be

Documentation in the United States Patent Office in the Field of Petroleum and Allied Subjects," Third World Petroleum Congress, The Hague, 1951. Proceedings, Section X. P. 13-21.

¹⁰ S. M. Newman, "Problems in Mechanizing the Search in Examining Patent Applications," (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.).

¹¹ Lanham, B. E., J. Leibowitz and H. R. Koller, "Advances in Mechanization of Patent Searching—Chemical Field." April 11, 1956. *J.P.O.S.*, 38, 820-838 (December, 1956). (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

¹² H. F. Lindenmeyer, "What does the Patent Office Scientific Library Have to Offer the Chemist?," *J.P.O.S.*, 36, 463-481 (July, 1954). (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

combined in separate documents as anticipatory, provided such type of combination is within the suggestion of one or more documents.

After the details concerning the scope and approach have been determined, they should be kept in mind continuously throughout the search while studying each document.

Unfortunately, no infallible or obvious procedure is always available to obtain prompt or ultimate identification of the specific subclass for each item to be searched. The complete search is not often limited to one or a group of subclasses or even to a single class, and it may thus be necessary to identify additional search fields for various phases of the subject matter sought.

In many cases, use of the alphabetical index of the Manual of Classification will help to locate the proper search areas quickly and accurately. If the term sought is not found, its synonyms should also be investigated. If the index is not used then the titles of the main classes should be scanned to select the one that appears pertinent. The classes in the Manual are given in numerical rather than subject order. The class titles usually indicate their relationship to the required search, but if there is any doubt other related classes should be compared. The definitions and notes of a class and its subclasses usually verify the search field.

After selecting the pertinent class, the titles of the major subclass groupings which appear in the first line of indentation should be read in sequential order. Once a title has been located which identifies the subject matter the coordinate major subclasses under it can usually be ignored. Minor and sub-minor subclasses indented under the selected major subclass should be investigated to determine if their titles also relate to the search required. Indented subclasses usually include species, their major subclass is generic, and includes species not provided for in its indentations. Almost every class includes a subclass bearing the heading "Miscellaneous"; this is a pigeon hole for patents which fall under the class definition, for which there is no specific subclass. Varying scopes of a given subject matter are not separately classified. The subclass numbers in many classes are not in numerical order and serve only for identification rather than superiority.

An example of the major and minor subclasses is illustrated below.

Class 260, Chemistry, Carbon Compounds, certain subclasses being listed in the following order:

- 239. Heterocyclic Carbon Compounds
- 298. Azoles
- 302. Thiazoles
- 303. Anthrone or anthraquinone nuclei
- 304. Arylenethiazoles
- 305. 2-amino
- 306. 2-thio

The full title of subclass 306, Class 260, is as follows: Chemistry, Carbon Compounds, Heterocyclic Carbon Compounds, Azoles, Thiazoles, Arylene-thiazoles, 2-thio. An attempt to locate the proper search area should definitely follow class plans and structures as well as definitions and notes, as created by various classifiers, since these factors differ among classes. If any difficulty or doubtfulness should arise in determining a proper field of search, the Patent Office will supply available information as to the identity of the pertinent classes and subclasses, provided the request includes specific details of the subject matter sought^{13, 14}.

Classification System Inadequacies

The shortcomings of Patent Office and other classification systems are to some extent similar¹⁵. The effectiveness of a system is dependent upon how closely the basis of its establishment is correlated with the basis of the required searches, but it is impossible for the patent classifier to provide for or even anticipate all the search viewpoints to be desired within the scope of a given class. Such problems detract from the efficiency of the searches made by the Examiner as well as by others. A few examples of such difficulties will be illustrated with respect to the chemical field.

When confronted with a search for a specific chemical compound, wherein all of the structural characteristics are set forth, the structural group present in the formula which appears highest in the subclass schedules will identify the precise field of search. Thiazole in Class 260 is an example. Generic searches, however, present a major problem. For example, if it should be desired to find disclosures of all compounds which contain an azole structure regardless of any other structures which may be attached thereto, the "azole" category in the schedule would not provide an adequate search. All superior categories may contain disclosures of the type sought but they have been classified on the basis of other fragments.

The search for generic processes involves the same problems. Those which result in a specific compound are classified therewith, but when the procedures are not limited to specific reactants and product the practical search field is not ordinarily identifiable.

Certain types of composition searches are also extremely difficult. Where

¹³ "Information Concerning Patent Classification and Patent Records." Two page circular. (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

¹⁴ B. E. Lanham, "Services Available from the Patent Office," *Special Libraries*, **46**, 25-28 (January, 1955). Elaboration on (13). (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

¹⁵ D. D. Andrews, "Modernizing Chemical Patent Classification," Presented at 128th ACS Meeting, Minneapolis, September 16, 1955. (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

the search is to determine if compound X has been disclosed as an ingredient in lubricants, and such compound is not included in the title or definition of an ingredient subclass, all patents in the 82 lubricant subclasses in Class 252, Compositions, must be investigated.

Another composition search problem is quite important, both to the Patent Office and inventors. A specific composition disclosed only for use as a detergent is not patentable over a previous disclosure of the same composition for another use, such as an adsorbent. The complete search for such a detergent composition, then, would possibly require investigation of all composition disclosures distributed among numerous classes.

No classification is based on adhesive compositions, and without a knowledge of the specific ingredients, an almost unlimited search in sections of several classes is required.

Another frequent approach to searching is by one who wishes to learn all of the uses of a material, such as titanium in alloys, compounds, manufactured articles, etc. No such search can be indicated since the desired information may be found in many of the classes.

Such difficulties cannot be avoided in the present classification system, nor is it practical from various viewpoints to revise and enlarge the system to a sufficient extent.

In 1946 the Patent Office considered the development of mechanized search procedures to provide facilities for searching any given subject matter from any required viewpoint.

The early experiments involved studies with respect to the use of edge-notched cards and the "unit card system." Since such methods were eventually determined to be impractical for Patent Office objectives they were abandoned in favor of what was considered a more feasible approach, the general description of which will be set forth later in this Chapter. The first will relate to the experiment which culminated in machine searching tests in the Spring of 1950; the second will deal with the current program which started subsequent to the recommendations of the Bush Committee¹⁶.

THE FIRST EXPERIMENT

Subject Matter

The first experiment was in the field of medicinal compositions, and a sample group consisting of 441 patents was selected to constitute the subject matter. The disclosures of these patents were considered to consist of two basic types of information units—ingredients and functions. The ingredients were chemical compounds, free elements and so-called "complexes," a term applied to materials described by language other than or

¹⁶ "Report to the Secretary of Commerce by the Advisory Committee on Application of Machines to Patent Office Operations," Washington, D. C., December 22, 1954.

in addition to the terminology of chemical compounds. Complexes included such things as "poppy seed oil," liver "extract," milk, etc.

The term "function" related to nontangible disclosures such as properties, uses, and behavior of materials. Such terms as "hormone," "lubricant," "malaria" were considered to be functions.

The Problem

A patent search involves, in effect, a question and an answer. The question may be formulated as follows. Is there a disclosure anywhere of the concept expressed in the claimed subject matter? The answer is provided by either finding or not finding such disclosure.

Several features are significant in patent searching. First, the searcher is initially unaware of the existence or non-existence of a disclosure; in contradistinction to a search where there is known to be a positive answer, such as a quest for the date of an historical event. Since a presumption that there is no reference to the disclosure is made on the basis of failure to find any, it is important that no pertinent detail, or place where such detail may be located, be overlooked.

Second, the search is based not on words but on the meaning or import of the words. Pertinency of disclosure relative to the subject matter is evaluated on the basis of the meaning of the disclosure as compared with the meaning of the subject matter of the claims.

Third, the search is often "generic," not only because claims may be presented in generic form but also because the Examiner may be searching for "related" subject matter to determine the question of invention in addition to the question of novelty. A generic search is met by finding any specific embodiment within the scope of the genus. It is not practically possible, however, for the searcher to envision all the members embodied by the genus and to thus express his search in terms of a collection of these specific members.

It will be evident that the search question and the disclosure which answers it will not ordinarily be in the same language or context. The subject matter of the search will very often be included within a more comprehensive disclosure context. Thus $A + B$ is included within $A + B + C$ and $A B$ is included with $A B C$; a search for a disclosure of mixture $A + B$ can be met by finding a disclosure of a mixture $A + B + C$ and a search for a chemical compound $A B$ can be met by a disclosure of a chemical compound $A B C$. Since the searcher is not generally aware of how the subject matter is disclosed he cannot be completely certain as to where it is classified and so may miss pertinent answers to his questions.

Since automatic data processing machines can sense and interpret fragments of a combination independently of each other, and of the combina-

tion, mechanization is expected to aid considerably in the solution of a search problem. If, for example, a given disclosure is analyzed into a relationship $A + B + C$, and this information is transcribed to the machine according to a logic which defines the relationship, the machine can be made to recognize A , B , $A + B$, and so on, each independently of the rest of the context. The problem from the point of view of mechanization is to select these units A , B , C , etc., to serve as "building blocks" and construct a logic for definition of the units and their relationships. These "building blocks," or "descriptors," as they may be called, would be recorded and manipulated by the machine according to a logic whereby the disclosure could be reconstituted in a manner congruent with the logic of the search requests.

Schedules of Descriptors

Pending the accumulation of sufficient experience to make an optimum selection of descriptors, the initial selection is done on a tentative and approximate basis. For chemical compounds the descriptors were of the same type already found useful for searching by conventional means. The functions were selected as a result of the analysis of the patent disclosures on the basis of educated guesses as to the terms most likely to be used in searching. A schedule of descriptor terminology was built up which consisted of four sections: the inorganic and organic sections, and the "complex" section which contained a classification of plants, animals, minerals, and terms deemed pertinent for identification and searching of the materials. The fourth section contained terms of function and included a classification of disease in terms of body systems and pathogenic organisms. The terms were generally arranged in order of decreasing genericity, indicated by indentation. Thus,

Heterocyclic Compounds	1313
Para-n-benzene Sulfoxy	1313-1512
Azoles	1313-2512
Thiazoles	1313-2512-1423
Oxazoles	1313-2512-1523

The codes reflected the same pattern of indentation. The most generic term of any particular aspect was called a "first position" term; the next indented term was a "second position" term and so on. The code of an indented descriptor contains the codes of all descriptors generic to it. Thus, "Thiazoles" contains the code for "Azoles" and for "Heterocyclic Compounds." Since the generic class descriptors were inherently within the more specific class descriptors a search by a generic term would retrieve disclosures of all materials falling within that class.

Each disclosure of a composition was arranged according to an organization of two levels, which may be symbolized as follows:

where the alphabetical symbols are the descriptors. The first level associated the descriptors within parentheses and the second level associated the parenthetical groups within brackets; similar to the association of letters in a word and words in a sentence. The collection of descriptors within each parenthesis represented an item and the collection of items within each bracket represented a mixture of items or composition.

The disclosure "olive oil, having a solvent function," can be represented as (E F G) where E is "a fatty acid ester," F is "an extract of the olive plant" and G represents a "solvent."

Thus, formula I above, exemplifies the organization for coding the following type disclosure "a mixture of sulfathiazole and olive oil, wherein sulfathiazole is a bacteriostatic, olive oil is a solvent and wherein the composition is for the treatment of tonsillitis."

The punched card used is illustrated in Figure 12-1. The first ten fields of the card were allotted for punching the document identification. Columns 13 to 80 inclusive were divided into seven sections corresponding to the seven positions of descriptors as they appeared on the schedule. A code represented one descriptor of the schedule and it was sectioned into its corresponding positions. Thus, for "thiazoles," a category in the 3rd position, the 1313-2512-1423, was punched in any horizontal row (shown in row 7 of Figure 12-1 in sections 1, 2 and 3, respectively. In scanning by machine, each horizontal row was sensed independently of any other row and the information was recognized as such regardless of the row in which it appeared. The codes pertaining to an item were associated with each other by a punch at the upper end of the group in column 12. The scanning

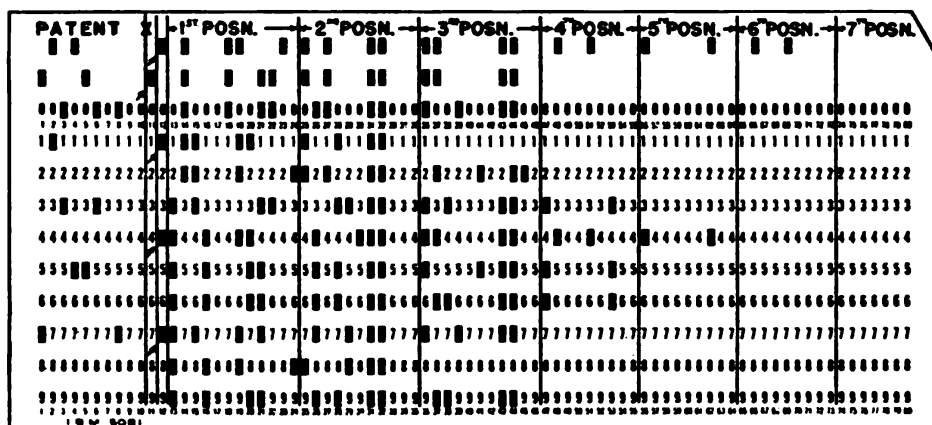


Figure 12-1.

was continuous from card to card and the association of items in a composition was signalled by the 2nd level punch in column 11. The last card of each group pertaining to a composition was selected if a hit had been registered, and the identifying data on the card indicated the source of the disclosure.

Searches

This coding organization permitted the finding of the disclosures in terms of any one or more of the descriptors regardless of the presence or absence of the other descriptors. The above hypothetical composition might be in conventional classification under sulfa drug medicines—Class 167, Medicines, Poisons and Cosmetics, subclass 51.5. A searcher for “an aromatic amine in admixture with a fatty acid ester” would find no obvious reason to search Class 167, subclass 51.5, yet said location would contain a fully pertinent reference for the desired search. A few examples of types of searches which would successfully retrieve the illustrated disclosures are:

- | | |
|-----------------|--|
| 1 (A) | A compound of the thiazole class |
| 2 (A C) | A thiazole—aromatic amine compound |
| 3 [(A) (G) (H)] | A thiazole plus a solvent for use in diseases of mucous membranes |
| 4 [(I)] | A composition for treatment of streptococcus infection regardless of what the ingredients are. |

Performance Tests

The tests were performed with respect to actual patent applications. The 441 patents containing 6,272 items described in terms of 18,650 descriptors were searched in 4.5 minutes or about 95 patents per minute. The princi-

ples described were successfully embodied. The card-sorting machine was a temporarily modified I.B.M., E.S.M. 101. Further details of this project have been previously published^{9, 17}.

THE SECOND EXPERIMENT

The first experiment indicated the feasibility of mechanizing the patent search and it was subsequently decided to expand to an operational basis for the entire chemical arts. In view of this broader program, new principles had to be developed to provide greater flexibility and effectiveness than was available according to the earlier project. Some of these developments, described as the second experiment, are being jointly undertaken with the National Bureau of Standards.

The description necessarily presents various segments of the over-all problem. Integration into a more unified and comprehensive picture is expected as the developments progress closer to completion. Related progress in the nonchemical arts has been described¹⁸.

The Descriptors

The descriptors used in the first experiment were generally of the "combinatory" type. For example:

1. Amine-containing compounds
2. With hydroxy groups
3. Aromatic

Category 2 does not involve more specific delineation of category 1 but involves instead a combination of category 1 with a group extraneous to it.

This type of schedule has certain advantages in that it sets forth relationships between different chemical groups. Thus category 3, by definition, sets forth a benzene structure containing both an amine group and a hydroxy group. The disadvantages, however, reside in the lack of genericity provided for categories 2 and 3. A search for hydroxy compounds or aromatic compounds, regardless of the presence or absence of other chemical groups, cannot be made according to this particular example.

The second system attempts to provide many more generic search aspects, while at the same time retaining the various relationships. In the

¹⁷ Bailey, M. F., B. E. Lanham and J. Leibowitz, "Mechanized Searching in the U. S. Patent Office," *J.P.O.S.*, 25, 566-587 (August 1953). (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

¹⁸ Andrews, D. D., and S. M. Newman. "Storage and Retrieval of Contents of Technical Literature, Nonchemical Information," May 15, 1956. (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

present schedules any indented descriptor refers to further specificity of the broader descriptor rather than a combination, as

Non-metal
Metal
 Light metal
 Heavy metal

The relationships among descriptors are shown by special devices which will be described.

Chemical Compound Coding

The structural formula representation of a chemical compound contains intrinsically the configuration of many chemical classes. Given a particular formula, the chemist can, on inspection, recognize any and all classes to which it belongs, insofar as these classes are definable in terms of an element configuration. By getting the machine to perform the same inspection and recognition, a compound would be available in terms of any structural class inherent within it without the need to preassign descriptors to it. The descriptors are thereby potentially available, in effect, to be synthesized as needed.

The meaning portrayed by the formula can be conveyed by a description of each element in the formula and its connectivity to other elements. This permits the finding of any compound in terms of any combination of elements in any structural arrangement within the molecule, independently of any other fragment or of the complete structure. Several methods for coding in this element by element fashion have been developed; determination of the optimum method is expected after adequate machine tests. The methods depend on what is called the "interfix" device.

The Interfix

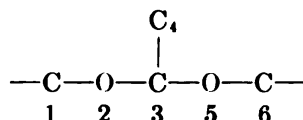
The interfixes are numerical descriptors wherein significance is in relative rather than absolute values. For example, in the fragment C^1O^2C connectivity is shown by sameness of interfix numbers, i.e., those elements which have the same interfix numbers are connected to each other. The fragment can be coded (C₁) (O₁O₂) (C₂). The same connectivity is also indicated by (C₅) (O_{5,9}) (C₉).

By this method, numbers are assigned to each connection between elements. The numbering may start at any point in the structural formula and is entirely random insofar as sequence is concerned. Elements which have the same number are connected, regardless of the absolute numerical value.

A variation of this "element by element" coding concept, which has

been successfully tested on the SEAC at the National Bureau of Standards involves the following method¹⁹.

The configuration

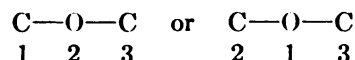


can be described (disregarding the types of bonds) by saying,

#1 C is connected to #2 O
 #2 O is connected to #1 C and #3 C
 #3 C is connected to #2 O, #5 O, and #4 C
 #4 C is connected to #3 C
 #5 O is connected to #3 C, #6 C
 #6 C is connected to #5 O

From this information the element configuration or any part thereof can be reconstructed. This would hold true no matter what numbers are assigned to the element, just so long as each element is uniquely identified.

For example



The same group is numbered two different ways. The first will be coded to mean

#1 C is connected to #2 O
 #2 O is connected to #1 C, #3 C
 #3 C is connected to #2 O

The second

#1 O is connected to #2 C, #3 C
 #2 C is connected to #1 O
 #3 C is connected to #1 O

Reconstruction of each of these different codes will yield the same structural fragment.

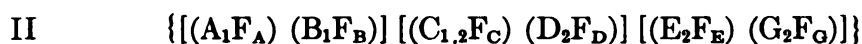
Another method for coding chemical compounds involves the use of two structural entities, the ring configuration and the chain configuration, to constitute two major building blocks. The connectivity of the elements in the ring as well as in the chain is shown by the coding sequence for each building block. Juncture among the blocks is indicated by two types of

¹⁹ Ray, L. C., and R. A. Kirsch, "The Use of Automatic Data Processing Systems in the Retrieval of Technical Information," National Bureau of Standards, a preliminary report. Publication is expected.

such structural interfix, i.e., the "shared element" interfix and the "bond juncture" interfix. The former applies to such joinings as exist in fused ring and spiro arrangements; the latter refers to attachments of chain to ring or ring to ring as in diphenyl. Additional descriptors are used to describe any configurations not intrinsic in any specific element arrangement, such as "amide," "acid," etc. More detailed accounts of this and other fragments of the over-all problem have been published¹¹.

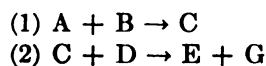
Organization of the Disclosure

Three levels of organization of the disclosure for coding are used.



The first and second levels are groupings for the item descriptors and the items, respectively, as in the first project. The third level indicated by the bowed brackets is the process level. For convenience in exposition a single alphabetical symbol is used to represent all the descriptors for each item, except for function. The function is indicated by F and the subscript to F indicates what material it is a function of.

The numbers are "sequence" interfixes. A higher number indicates a later step in a reaction. Thus formula II, in terms of process, may be read as follows:



In addition to the process, the formula may indicate three separate compositions, each containing two items. Thus by the interfix, relationships can be shown which cut across the groupings. C and D are in the same group as ingredients of a composition, but in different groups insofar as process steps involving the composition are concerned. The indications as to which are starting materials and which are results of the process are shown by descriptors associated with the ingredients.

A process, "N1 P1 Q2 R3," symbolizes that N and P are carried on simultaneously, and that reaction N and P each precede Q and R.

The function never appears as a separate item but is always on the first level associated with the ingredient. Where the function is related to a particular composition it is distributed as a descriptor to each ingredient of the composition possessing said function. Thus, if A plus B is an insecticidal mixture and on the addition of C, the insecticidal property is no longer existent but the new admixture functions as a herbicide, this disclosure is symbolized in formula III.



where I is "insecticide" and H is "herbicide" and "p" indicates a partial or shared function, which function is shared with any other ingredient having the same function descriptor.

Alternatives

Many disclosures of the so-called alternative type are found in patents. For example, "A is mixed with B or C". Several relationships are thereby shown.

- (1) A + B
- (2) A + C
- (3) C is alternative to B for use in admixture with A

Not shown is

- (1) B + C
- (2) A + B + C

The disclosure should not be selected in a search for B + C. Another "al-

ternative" situation is found in chemical formulas of the type R where



X is one of a, b, c and Y is one of d, e, etc.

The searcher cannot, of course, have any foreknowledge of the existence of an alternative situation with respect to the particular combination in which he is interested. He cannot specify, as a practical possibility, "A + B but not if A is alternative to B" since the alternativeness may exist on any level with respect to any element or group of elements. The "alternative" situation is therefore handled by a special signal, grouping the members of any alternative group, wherein an automatic discrimination is made for the proper selection according to the logical rules embracing the alternative situation.

Modulants

The modulants are ways of getting more versatility from the schedules and showing many more relationships. For example, the following is a set of modulants.

Disease of	50
Disease by	60
Infection by	60-10
Toxicity by	60-20
Ingredient	70

If the term for a particular plant such as a mold is associated with code 60-10, the term has been modulated to express the idea of a disease by infection with that mold. If code 50 is associated with the mold, it is thereby indicated that the mold itself is diseased. Code 70 associated with the mold indicates the mold to be an ingredient in a composition. Code 60-20 associated with a particular material expresses the idea of a disease of poisoning by that material. Thus, the modulants are used to inflect or vary the meaning of the root terms on the schedule.

Negatives

Certain disclosures are expressed in negative form such as "phenols may be used except those containing halogens." Provision has been made for finding the disclosure of positive assertions as to the absence of chemical groups.

SEAC TEST

A test of the structural formula search method has been made, as previously indicated, using the SEAC at the Bureau of Standards, and preparations are being made to test the logic of the comprehensive system involving the entire chemical field.^{20,21} The computer will be used to simulate a searching machine and it is expected, as a result of these tests, to evaluate the principles and logic of the system and to determine the requirements of an optimum search machine.

²⁰ B. E. Lanhan, J. Leibowitz, H. R. Koller and H. Pfeffer, "Organization of Chemical Disclosures for Mechanized Retrieval." Presented at 131st ACS Meeting, Miami, Florida, April 8, 1957. (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

²¹ H. Pfeffer, H. R. Koller and E. Marden, "A First Approach to the Patent Search on a Digital Computer (SEAC)." Presented at the 12th National Meeting of the Association for Computing Machinery, Houston, Texas, June 20, 1957. (Copies are available from Research and Development, U. S. Patent Office, Washington 25, D. C.)

Chapter 13

APPLICATION OF PUNCHED CARDS TO LIBRARY ROUTINES

MADELINE M. BERRY

National Science Foundation, Washington, D. C.

Introduction

Punched cards are a convenient tool for the recording of index entries to scientific and technical documents, and for the finding of references to pertinent documents in answer to given questions. The use of punched cards for such operations is described in other chapters in this book. These tools can also be applied to other tasks, such as the many clerical and technical routines necessary for the functioning of a library. The use of punched-card systems offers the possibility of performing clerical routines with a minimum of time and effort, of relieving the drudgery often associated with these tasks, and of freeing staff time to be devoted profitably to more professional work.

This chapter presents a summary of actual and possible applications of punched-card systems to library routines. There is no attempt here at exhaustive coverage. Nor is it possible to describe the systems in sufficient detail to enable installation of a system in a given situation. Rather the chapter tries to suggest ways in which punched-card systems, manual or mechanical, can be applied to advantage in performing many library operations. The reader is referred to "Library Applications of Punched Cards: A Description of Mechanical Systems" by Parker¹ and "Marginal Punched Cards in College and Research Libraries" by McGaw², as well as to other texts and articles noted, for more complete discussion of the subject.

Library routines may be divided into processing functions and reference functions. As noted above, punched-card systems have been used successfully in literature reference work. Other chapters of this book describe such applications, at least in principle. This chapter will consider only the processing functions in libraries. These functions may be considered to include those steps involved in preparing books for use—ordering and acquisition, binding and cataloging; and those involved in the use of books—

¹ Parker, Ralph H., *Library Applications of Punched Cards: A Description of Mechanical Systems*, Chicago, American Library Association, 1952.

² McGaw, Howard F., *Marginal Punched Cards in College and Research Libraries*, Washington 7, D. C., Scarecrow Press, 1952.

circulation. Personnel and financial administration routines in libraries can be performed by punched-card systems as in industrial organizations, so they will not be considered here.

Ordering and Acquisition

For the performance of *ordering* routines, punched cards have been used to replace the purchase order form typed and filed in multiple copies. The cards, manual or mechanically-sorted, may show any or all of these items of information: author of the volume ordered, its title, dealer or agent through whom purchases are made, and date ordered. In mechanically-sorted systems especially, the cards may also show estimated price, order number, and such information. When books are received, they are checked against the "on order" file and the cards are punched or notched with the date received and perhaps the actual price. This file then can be thought of as the master record of volumes purchased by the library. If the cards contain additional data, such as subject classification of the book, age group to which it is applicable, the language of the book, its literary form (i.e., fiction or poetry), or the country of its origin, this file may help the library administrative staff in the molding of *acquisition policy*. The Montclair Public Library, in Montclair, New Jersey, uses its IBM-card file to make such analyses for policy decisions (Figure 13-1). Thus types of non-fiction bought for adults, book purchases made for children, sources of purchases, languages represented, and additions to special collections have all been listed easily by proper sorting of the punched cards. It has been possible, for example, to learn the extent of use and popularity of material according to the date of publication³.

Various forms of cards have been used successfully for ordering and acquisition work. An IBM card system is used in the Order Department of the University of Florida Libraries⁴. The Boston Public Library uses a dual-portion IBM card for purchasing records, and prepares purchase analyses reports, including accounts of net value for various funds, reports of the status of the City Fund, and the like. When the cards have been used for such tabulations for a particular period, they are sorted by title and author and filed in the Purchasing Department for further reference⁵.

³ Quigley, Margery, "Business Machines in a Public Library," *American City*, 60, 101-2 (1945).

Numerous personal communications and unpublished reports were also made available by Miss Quigley.

⁴ Duer, Margaret M., and Lewis, Clark S., "How We Use IBM," *Library Journal*, 78, 1288-9 (1953).

⁵ "Purchase Analysis Procedure—Boston Public Library" (mimeographed) New York, International Business Machines Corp., 1934.

Figure 13-1. Punched card used at Montclair Public Library.

The IBM punched-card installation at the Milwaukee Public Library⁶ has proved its value in *book budget accounting*. The file furnishes monthly cumulative totals of money paid out to dealers, of orders still outstanding, and of the total money remaining in the book budget. This operation aids in spreading expenditures and the work of processing books received. It also allows a consistent follow-up on items not received. Easy access to the information on book purchases, as supplied by the Tabulating Division, guides the library staff in its purchasing and discard policies. Studies are made of the elapsed time between order and receipt of a book, so techniques of purchasing can be improved and the flow of work better organized. The punched-card file can even be used to handle payments of book and periodical invoices. Other sets of cards are used in *shelf-listing*. The breakdown of subject matter according to major classification divisions (by Dewey numbers) shows where holdings are inadequate, or out-of-proportion, as an aid to buying or discarding procedures.

Some of the advantages gained by adoption of mechanically-sorted punched-card systems can also be obtained by use of manually-sorted cards. One consideration in the use of mechanical equipment is the expense of its installation and operation *versus* the amount of work to be done. One merit of the manual systems for small installations lies in the fact that the only special equipment required is a hand punch and a sorting needle. The University of Illinois Library⁷ uses McBee Keysort cards for its *acquisition records*. An order card (Figure 13-2) is made for each title to be purchased. The fund on which the book is to be purchased, the agent,

⁶ Baatz, Silmer H., and Maurer, Eugene H., "Machines at Work," *Library Journal*, 78, 1277-81 (1953).

⁷ Brown, G. B., "Use of Punch Cards in Acquisition Work: Experience at Illinois," *College and Research Libraries*, 10, 219-20 (July 1949).

1 2 3 4				PART COM				7 4 2 1				7 4 2 1				7 4 2 1			
DATE WANTED IF RUSH				AUTHOR												DATE ORDERED			
DATE ORDERED				TITLE												OF			
COST				SERIES												DATE OF BILL			
CHARGED TO				EDITION				PLACE				STAMP LOCATION				CIRC. CHARGE			
				PUBLISHED								GEN. LIB. STACKS				OFFICE COLLECTION			
				DATE PUB.				VOL. I.				NO. OF COPIES				PRICE			
				FUND				RECOMMENDED BY				APPROVED BY							
UNIV. OF ILL. LIB.				DATE FWD.				CITE IN LOG. REFERENCE ON VERSO											

Figure 13-2. Keysort card used for acquisition record at University of Illinois Library.

and the purchase order number are recorded on the card. In addition, the codes for author and fund are punched in. When the book and invoice are received, the card is coded complete or partially complete depending on whether all volumes of the title have been supplied, and the card is refiled in the orders and receipts file until the book is cataloged. At that time, the year of receipt is coded in the upper left-hand corner and the card is placed in the dead file, from which it is discarded after a period of three years. The system shows advantages in speeding up the processing of invoices, allowing efficient follow-up of orders, and making the filing of cards a fast and almost mechanical task. One complete step, that of checking off the invoice against the copy of the purchase order, is eliminated. Instead, as soon as the invoice is approved, it is entered as a disbursement in the ledger and paid. At quarterly intervals all outstanding orders are sorted by fund. Encumbrances on each fund can then be corrected and overdue orders are claimed.

In many libraries the staff preferred to use multiple slip order forms, filed in several ways. It is now possible to get manual punched cards in multiple slip form to provide the advantages of both approaches.

Serials ordering procedures can also be facilitated by the application of punched cards. Frequent sorting of the file of cards enables catching subscriptions before they lapse. At Pennsylvania State University, a marginal punched-card system has permitted great savings in time when sorting for

order number, and country, and other special tabulations are made as requested⁸.

Punched cards are also used for fiscal controls of book funds. Semi-monthly listings are made for the two major appropriations by form of order (Regular, Blanket, and Continuation); within each type of order a further breakdown is made by Recommending Officer or form of material sub-allotment. A yearly report is also made for pieces purchased by fund, country, and form of material. It serves as a summation of all serial publications purchased in the year.

Cards for serial-ordering routines usually have space for coding the country of origin, frequency of publication, language, price, renewal date, source, subject, type, and so forth. The Milwaukee Public Library uses its IBM card files to analyze serial purchases as it does book orders: who gets what, and the strengths or weaknesses in coverage of the various subject areas. Such analyses aid in determining purchasing or cancellation policies.

The Library at Dow Chemical Company in Midland, Michigan, uses an IBM card system for renewal of serial subscriptions. One set of cards has titles of journals, another set has addresses of vendors through whom the journals are ordered. The two sets are tied together with the identifying numbers assigned to each serial publication. When the various departments of the company have checked the list of journals they subscribe to, the address file is rearranged by vendor to facilitate the renewal procedure. The purchase order then takes the form of lists of titles prepared from the punched cards and sent to the appropriate vendor⁹.

Punched cards are also a valuable tool in keeping records for *inventory control*, and for measuring a library's resources. The value of any missing books can be calculated, or the monetary value and number of all the books in a given subject class can be computed. The number of duplicate copies of documents, the number of gift volumes, the number of volumes in a given language, can all be listed.

The IBM card inventory deck at the Dow Chemical Company Library is used to determine *shelf space requirements* and rate of growth of the library's collection. The file is checked every three years to determine optimum spacing on the shelves for the next period⁹.

⁸ Keller, Alton H., "Book Records on Punched Cards," *Library Journal*, 71, 1785-6 (Dec. 15, 1946).

Personnel of the Library of Congress gave invaluable help in preparing up-to-date descriptions of the use of punched cards at the Library. Their assistance is appreciated.

⁹ Taylor, F. Lowell, personal communication, June 1956.

Binding

Better control of material from the time of ordering until final discarding can be provided by punched-card files. Thus *binding schedules* or bindery records can be maintained by means of such systems. The cards might contain codes for the month and year of publication of the material gathered for binding, the title and call number, the type of publication, color of binding, the size, the number of volumes to be bound as one, and such pertinent data. The University of Georgia uses a McBee Keysort system for these records², as does the library at Dow Chemical Company. On the other hand, a mechanized system can be used for the same routine, with the possibility of searching each month for titles scheduled to be gathered together and sent to the bindery. It is possible to include the information about binding with the information about subscriptions on a single card¹.

Cataloging

The process of *cataloging* may also be facilitated by use of punched cards. This is especially true in the matter of producing *catalog lists* by means of machine-sorted cards. For example, the King County Public Library in Seattle assembles and prints its catalog by the IBM system¹⁰. The library has many branches, with constantly changing catalogs. The mechanized system relieves the drudgery and time-consuming task of removing cards for books sent back to the main library and of adding cards for books which have been newly acquired. Instead, lists of names of books are tabulated with the IBM equipment at the main library, and assembled in looseleaf style. The lists are arranged as a catalog of adult books, one of juvenile books, and a combined alphabetic list, plus an author catalog. The lists are changed about every six weeks. The master cards from which the various lists are prepared include a classification number and the symbol used with it (e.g., B = bibliography, J = children's book), author, title, language, reading level, and a code for the subject matter. A duplicate file, called the locator file, is maintained. When books are sent out to the branches, the card is stamped with the branch name and the date it was sent out.

The simplified subject codes, punched in the master cards as mentioned above, are arranged alphabetically and assigned consecutive numbers. Spaces are left between the code numbers to allow for additional subjects to be included in the alphabetical list. Since these codes are general, a key or index is prepared to direct the user to the heading under which to look for a given subject (e.g., accounting, household—under home economics).

¹⁰ Alvord, Dorothy, "King County Public Library Does It With IBM," *Pacific Northwest Library Association Quart.*, April 1952.

The circulation record cards at the Milwaukee Public Library are used in like manner to prepare lists used as catalogs on the Bookmobile⁶.

The Library of Congress prepares a continuing supplement to the *Union List of Serials* as a monthly listing called *New Serial Titles*, which is available on a subscription basis. Punched-card methods are used in the preparation of these lists. Serials first published after December 31, 1949, and received by the Library of Congress and nearly 300 cooperating libraries are arranged alphabetically. The listing (Figure 13-4) shows title of the

660
CHEMICAL INDUSTRY AND ENGINEERING. SYDNEY.
1, MY 1953-
V. 1, NO. 3-5, JL-S 1953 OUT OF PRINT.
MONTHLY. E.G. HOLT PUBLISHERS, 166 PHILIP
STREET, SIDNEY, AUSTRALIA. \$3.00
K U 4- N N 1-

ENC 5600027660		22001	NN 1-
ENC 5600027660		22001	K U 4-
ENC 5600027660		22001	STREET, SYDNEY, AUSTRALIA. \$3.00
ENC 5600027660		22001	MONTHLY. E.G. HOLT PUBLISHERS, 166 PHILIP
ENC 5600027660		22001	4. 1. NO. 3-5. JUL 5 1953 OUT OF PRINT.
ENC 5600027660		22001	1. JAN 1953
ENC 5600027660		22001	CHEMICAL INDUSTRY AND ENGINEERING. SYDNEY.
ENC 5600027660		22001	BIBLIOGRAPHIC DESCRIPTION, NOTES, HOLDINGS, ETC.

Figure 13-4. Listing (above) and punched cards (below) used in preparing the *Union List of Serials* at the Library of Congress.

serial, place of publication, frequency of issue, the holding libraries (by National Union Catalog symbols) and the issue with which the library's holdings began. The entries are coded on the punched cards by subject content, language, and country of origin, so it is possible to make general listings by subject, country or language arrangements, or special listings of serials on given subjects, from given countries, or in given languages. Lists in classed subject arrangement, for example, appear in twelve monthly issues, sold on a subscription basis like the alphabetic list.

The alphabetic lists appear monthly and in annual cumulation which are self-cumulative over five-year periods. Using punched cards for the entries makes such annual and five-year cumulations easier to prepare and to print.

Another interesting project is one now going forward in Italy, in which a national Union Catalog of Italian libraries is being prepared by means of Remington Rand punched cards. Thirteen important libraries are recording their holdings on cards or tapes, which are then collected at the National Central Library Vittorio Emanuel. The tapes are converted to cards, and then the cards are alphabetized, checked for duplicates, and a single master file produced. From this deck will be produced sets of cards or sheets of paper printed with the bibliographic information. These cards or lists will be sent to all the most important Italian libraries, which will then check their holdings against the holding of the Roman libraries. Adding to and revising the first lists will eventually result in a Union Catalog for all Italian libraries.

Additional ways in which to use the information recorded on such cards in preparation of catalogs cannot always be predicted in advance. Such information might permit compilation of statistics or the conducting of studies which may be necessary for justification of budgets, for definition of policy, or for public relations work.

Circulation

Keeping track of the actual use of books by maintaining circulation records is the work area with the greatest potential for the application of punched-card systems. Both manual and mechanically-sorted systems are much in evidence in public libraries, those of colleges and universities, and special libraries. The use of punched cards helps reduce the clerical work necessary in arranging files of cards, stamping or writing identification numbers and dates, or sending overdue notices. Basically, circulation entails transcribing the borrower's identification to a book card, recording the date borrowed or the date due, and filing the book card by date for retrieval when the book is returned. Ordinarily this is the only kind of record kept by the public libraries. College and university libraries, because of their reference function and the consequent need for more flexible con-

trols, usually maintain class records and borrowers' records in addition to the date files. Punched cards make it possible to combine more than one such file into a single master file. This gives only one place to search for a charge. Some libraries keep two files—one for active books and one for inactive books. This still means that there are only two places in which to make a complete search, and usually it is necessary to consult only one of the files.

Charging systems for circulation control can involve three types of records: transaction cards, call cards, and book cards. Any of these may be in the form of punched cards, with consequent advantages and limitations.

A charging system that uses punched-card transaction cards necessitates the following procedure. A call card or slip is filled out by the borrower and stamped with the transaction number and date it is borrowed. A pre-punched, pre-dated and numbered transaction card corresponding to the call card is slipped into the book pocket. When the book is returned, the transaction card is removed and the book is ready for circulation. The transaction cards can be sorted by month and day, and those representing books not yet due can be filed by date. On the due date, all cards received as books are returned can be sorted by number and matched against a master deck. Missing numbers represent overdue books. Call slips with the same numbers can be pulled for the borrowers' names, and overdue notices sent out. The advantages of such a system are that it is speedy, accurate and efficient. One card provides information about the book and the date due and only one card is used for the charging procedure. There is less manual labor involved and therefore fewer library assistants are needed for this particular job. On the other hand, the location of a specific book is difficult to trace. The borrower has no record of what he has charged out, and there is no recorded proof of a book having been returned. Another disadvantage is that it is difficult to take inventory of the library's holdings at any given time.

The Detroit Public Library maintains such a charging system on IBM cards¹¹. Loan slips are stamped numerically with serial loan numbers, the date due, agency (or branch library) name and its identification number. The borrower merely signs his name and address and the call number of the book he wishes to take. A punched card, with the same information punched into it as is stamped on the loan slip, is inserted into the book pocket as the date card. When the book is returned, this card is removed and the book is again ready for circulation. Such a system eliminates any delay in finding the book card and re-inserting it before returning the book to circulation.

¹¹ Monkevich, Edward, "Public Library Mechanizes Book Loans," *The Punched Card*, 1, 140-2 (1952-53).

Punched card sets are prepared in advance for each agency (branch library) with the year, due date, branch identification number, transaction number and deck (set) number. The cards from returned books are sent into the tabulating room daily. The "overdue" cards are sorted out, and the rest of the deck set aside until the due date is passed. The cards are then sorted, damaged cards are replaced, the deck is compared with a master deck and missing cards added, a new due date is punched in and the set is then returned to the proper branch. The missing cards, of course, represent books which are overdue, and overdue notices are typed and mailed to the borrowers.

The system assigns a fixed day each week as the date due for each agency or branch. The books are circulated for four weeks, with this fixed due date. This practice simplified the charge file and reduced the number of overdue routines. The charging system in general is a speedier, more accurate, and more businesslike method for controlling circulation.

The Free Library of Philadelphia has installed the same type of charging system using transaction cards, but with two important modifications. The charge-out step is done by a photographic method. Transaction card, book card, and borrower's card are placed together in a Diebold Flofilmer camera and microfilmed. Thus at the end of each day, the library has a film record of its transactions, with complete details about the book borrowed and the identification of the borrower. The allowed loan period is the same for all books. When a due date has passed, the returned transaction cards are checked for missing numbers which represent overdue books. The microfilm is reviewed for the names of the delinquent borrowers as well as the names of the books. The overdue operation for the entire Philadelphia library system is located at the central library. Three clerks can handle overdues for its 39 branches. The use of the microfilm charging method makes it possible to eliminate stamping and writing on book cards and borrowers' cards, thus increasing the speed and accuracy of the operation.

The other modification employed at Philadelphia is the use of small 40-column punched cards as transaction cards (Figure 13-5). These cards, made by Underwood Corporation Samas Punched Card Division, measure approximately $2 \times 4\frac{3}{4}$ inches. Each card bears two due dates so it can be reused at six-month intervals. The new procedures using punched transaction cards and film charging are said to have released over twenty-five trained librarians from tedious clerical work for professional librarian duties¹².

The Brooklyn College Library also uses a punched transaction card sys-

¹² "The Philadelphia Story" and "The Free Library of Philadelphia Has More Efficient Book Control With Punched Cards," brochures published by Diebold Inc. and Underwood Corp. respectively.

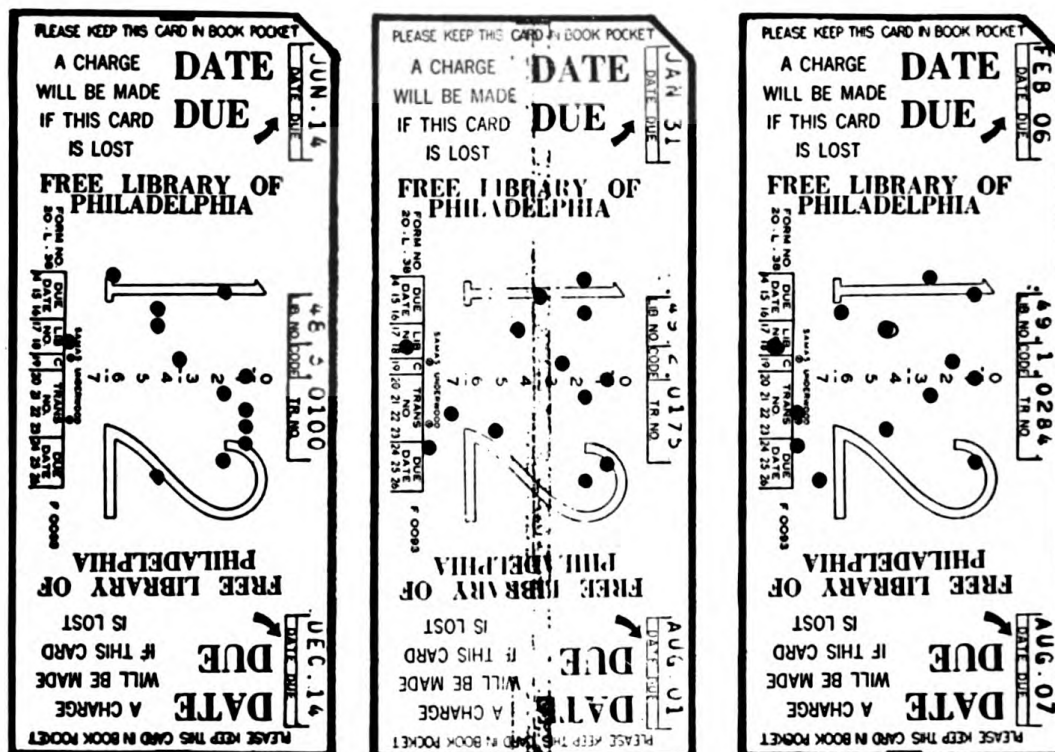


Figure 13-5. Transaction cards used at the Free Library of Philadelphia.

tem, with IBM cards. Sets of 600 consecutively-numbered cards are used, one set for each due date, with that due date stamped on and punched in. The operations of sorting the cards from returned books, collating them against a master deck to find missing cards for overdue books, and sending out overdue notices are the same as described above¹³.

Call cards can also be punched cards, of either the manual or machine-sorted type. In a charging procedure, the call card is made out by the borrower, and the date due is stamped on the call card and on a slip in the book pocket. The call cards are sorted according to the date due, and are punched or notched for that date. These cards can then be filed manually by classification number. When a book is returned the call card is removed from the file. Periodically the rest of the file is sorted by due date and overdue notices are sent out. With manual cards of the McBee Key-sort variety the notches for due date can be covered up and a new date can be punched, corresponding to a week later, to call attention to books still overdue, requiring second notices. The advantages of such a system

¹³ "Recruiting Library Personnel. Automation in the Library," 41st Conference of Eastern College Libraries, Columbia University, November 26, 1955. Published as ACRL Monograph 17, Chicago, Association of College and Reference Libraries, 1956, p. 34.

DO NOT REMOVE FROM BOOK POCKET	FILL OUT COMPLETELY; PRESENT AT CIRCULATION DESK	
	IF BOOK IS NOT AVAILABLE REASON FOR NON-DELIVERY WILL BE NOTED BELOW:	CALL NUMBER
	☐ ON LOAN; DUE _____	AUTHOR
	☐ WILL BE RECALLED IF REQUESTED	TITLE
	☐ AT BINBERTY	VOLUME
	☐ IN: _____	COPY
	☐ MISSING	CHECK ☐ FACULTY ☐ GRADUATE STUDENT ☐ UNDERGRADUATE ☐ NON-STUDENT
	☐ NOT LOCATED; FURTHER SEARCH WILL BE MADE. IF YOU WISH, YOU WILL BE NOTIFIED.	TEL. NO.
UNIVERSITY OF MISSOURI LIBRARY FORM 18-0	SIGNATURE	LOCAL ADDRESS

Figure 13-6. Punched card used as call slip at University of Missouri.

are its speed, accuracy and efficiency. One card gives a record of the book and the date due. Overdues can be handled easily by simple sorting procedures. Conversion from other systems is not difficult and the records are simple and flexible. One disadvantage of the procedure is that punched-card call cards are more expensive than the usual call slips.

The University of Wisconsin maintains its circulation records by using IBM cards as call slips. These cards are filled out by the borrowers, and when the book is taken the cards are kept at the circulation desk until the end of the day. The date due is then gang-punched through all the cards. They are then interfiled manually. Twice a week, the cards are sorted and those which represent books overdue drop out. Notices are then sent out manually¹⁴.

The University of Missouri also uses IBM cards as call slips (Figure 13-6). The original slip is placed in the card pocket of the book and is used to discharge the circulation record when the book is returned. Overdue notices are sent out weekly by making a Thermofax copy of the IBM card in the file.

Book cards in the form of punched cards are used in the same manner as conventional book cards. When the card is a manual punched card such as a McBee card, the date due and other information can be notched into it. The book card is signed by the borrower and a date due card is slipped into the pocket of the book. The book cards are sorted and filed by classification number. If not already so recorded the due date is punched or notched in. When a book is returned, the book card is put back in the pocket. Remaining cards can then be sorted by due date in order to send out overdue notices. The cards can be repunched for later notices, as described above. Such a charging system has all the advantages of the sys-

¹⁴ *Ibid.*, pp. 33-44.

tems described above. One disadvantage of these systems, when manual punched cards are used, is the necessity of plugging up notches before recording new dates such as required for additional overdue notices. However, all these systems use punched cards to reduce to a single file the number of circulation records usually maintained.

The Montclair Public Library has one of the country's most highly publicized mechanized circulation control systems³. It has been described in various publications, so it will merely be summarized here. Each IBM card represents one book or one borrower (Figure 13-7). Both types of cards, borrower's identification card and book card, are made from appropriate portions of master cards (Figure 13-1). These master cards have the desired information both written in and punched in. The borrower's card carries the person's identification number, address, voting district, age, sex, education, and occupation. The book card on the other hand has the

Record of book (Top Card):

ROBINSON THE CARDINAL
 20 7981
 CLASS NUMBER
 ACCESSION NUMBER
 LOCATION (BRANCH OR OTHER)
 DATE OF PUBLICATION
 COST
 1109300 62

Borrower's card (Bottom Card):

RUSH MRS HOWARD C
 19 WALNUT ST 81301
 BORROWER'S NUMBER
 1111111111
 DATE
 1111111111

Figure 13.7. Punched cards used as book card (above) borrower's identification card (below) at Montclair Public Library.

mailing to the borrowers. Renewals can be made by phone, in which case the book accession number is used to pull the loan card from the deck which is sorted and filed by these numbers. Then new loan cards are made with the new due date recorded. Reserves can also be handled easily with this punch-card system. A card is separated into two parts, one of which is filed at the loan desk for manual checking as books are returned. The other part of the card is used to make a duplicate loan card which can be collated with the deck of return cards to locate any returned books which have been reserved.

Fines and fees are also recorded by the control machine. When the transaction cards are sorted daily, those on which fines or fees have been paid are thrown out and a tabulation is made of the money collected for the day.

Uncataloged material also may be handled by the punched card system. The card which corresponds to the book card is pre-numbered and pre-punched with the number. The borrower writes down the title of the material and then one part of the card is used as a temporary book card and handled as described previously for the preparation of a loan card.

The system has been well received at Montclair Public Library by staff members as well as by the public. It shows advantages of speed, since the recording of loans and returns is accomplished by simply pressing a lever at the control machine; of accuracy, since numbers are automatically verified and charging errors eliminated; and of rapid turnover of books, since the book card is in the pocket at all times and the book is available immediately upon being returned. In addition, circulation statistics are accurately recorded and accumulated, and stored for further analysis and study.

Some libraries with special circulation problems have found punched cards to be a useful tool in maintaining control of material. For example, the Division for the Blind at the Library of Congress keeps track, by means of an IBM card system, of its 30,000 to 40,000 talking book machines. Approximately 6,000 to 8,000 new machines are purchased each year to replace worn or out-of-date models. Fifty-four distributing agencies for the machines send in records of loans, transfer of custody, repairs, or other status of the machines to the Library (Figure 13-9). The punched-card system makes possible the mechanical listing of all machines charged to each agency, as an annual inventory record. Each agency checks this list against its own records and sends any corrections, additions, or other notations to the Library, where the punched cards are changed accordingly. Listings can also be made for salvaged and discarded machines, arranged by model and serial number. Analysis of the listings reveals, for example, areas with excessive storage of machines, indicating that reader demand is not heavy enough to justify additional machines or that some machines registered there can be transferred to busier areas.

Figure 13-9. Punched cards for maintaining control of talking book machines at Library of Congress.

The Loan Division of the Library of Congress has another unusual use for punched cards. They are used to prepare lists for recall of material borrowed by the libraries of various Government agencies in Washington. The cards contain L. C. classification number, author, title, and place of imprinting of the material borrowed, plus the borrower's code number and the date the material was borrowed (Figure 13-10). These cards are pre-

DK565 F7										FRITSCH THE ALGAE VI CAMP 1935										AAS 05245705500									
CLASSIFICATION NUMBER										AUTHOR AND TITLE										IMPRINT VOLUMES DATE BORROWER									
DATE										CLASSIFICATION NUMBER										IMPRINT VOLUMES DATE BORROWER									
AUTHOR										TITLE										IMPRINT VOLUMES									
BORROWER										IMPRINT VOLUMES										IMPRINT VOLUMES									

LIBRARY OF CONGRESS
LOAN DIVISION

IBM 85727

Figure 13-10. Punched card used for recall of material borrowed from the Library of Congress.

pared at the Loan Division office for each book before it goes out on loan. Four copies are filed manually: one as a shelf-list; one by the borrower's code, sub-filed chronologically; one chronologically; and the fourth in the Library's central charge file. Before being sent to the central file the latter are cut down in size to eliminate the borrower's code punching. This retains as confidential the information about who borrows what. The file arranged chronologically has proved valuable as a tool for analysis of acquisition and discard policies. The Loan Division states, however, that the IBM card system was developed and is maintained primarily because of the ease of preparing recall lists. Since about 100,000 loans are made per year, sending out overdue notices is a tremendous task. Every three weeks the cards remaining in the borrowers' file in the subsection for the current return date are pulled and tabulated in separate lists, one for each borrower. These lists then are merely slipped into envelopes and mailed out.

Many libraries, both large and small, use hand-sorted punched-card systems for circulation work. As stated above, punched cards obviate the necessity of maintaining more than one file. For example, two files, one of book cards arranged by call number, and one of call slips (made out by borrowers) arranged by due date, can be combined in one file arranged by call number but notched by due date for easy sorting. Thus the cards can be easily pulled when books are returned, and needed for books which are overdue. Some systems use small edge-punched cards with holes for due dates, such as three days for each of a number of successive weeks (three, four or five weeks). Direct punching records the date due. The cards are sorted on three days, the cards drop out for books not yet returned, and overdue notices are sent out. The cards are punched for the same day in the next week, so that they will drop out then if the books have still not

MODEL, ATHERS, G. B. GREENSBORO C21594C

CALL NUMBER

← PLEASE PRINT

NUMBER FROM UPPER LEFT CORNER OF THE CATALOG CARD IN THIS SPACE. FICTION HAS NO CALL NUMBER.

VOL. _____

COPY _____ CHECKER _____

AUTHOR _____

BRIEF TITLE _____

SIGNATURE _____

ADDRESS _____

CITY _____ TELEPHONE _____

CHECK (X) BELOW

UNDER GRAD.	GRAD. ST'DNT	SPEC. STUDENT	GRAD. ASST.	FAC. ULTY	LIBRY. STAFF	OTHER
☐	☐	☐	☐	☐	☐	☐

UNIVERSITY OF NORTH CAROLINA LIBRARY

WEEK A			WEEK B			WEEK C		
MON	WED	FRI	MON	WED	FRI	MON	WED	FRI
☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 13-11. Call cards used for control purposes at University of North Carolina Library.

been returned. The needling, sorting and refileing of cards are accomplished in short order. One example² shows 50,000 cards sorted by a staff assistant in 1½ hours. Arranging such a file by first, second, and third notices takes fifteen minutes; rearranging and refileing by call number requires one-half to one hour.

The University of North Carolina Library uses colored clips to indicate cards which have been sorted as overdue and for which notices have been sent. There is then no need to sort for the relatively few overdues requiring second notices. The cards used in this system are not perforated at the top (Figure 13-11), so the clips don't get in the way of the holes or interfere with sorting operations. A smooth, flat metal clip must be used to prevent cards from catching on other cards^{2, 15}. The clips are also used on cards for

¹⁵ Hood, M., and Lyle, G. R., "New System of Book Charging for College Libraries," *Library Journal*, 65, 18-20 (Jan. 1940).

LOAN NUMBER

KEEP THIS CARD
IN THIS POCKET

BOOK IS DUE ON LAST DATE STAMPED

DATE DUE

DATE DUE

LOAN NUMBER

WAYNE COUNTY LIBRARY

BRANCH NUMBER

PLAIN
PINK
RED
BLACK
GREEN
YELLOW
BROWN
BLUE

Figure 13-12. Card used for circulation control at Wayne County Public Library.

books to be held for other borrowers and for books currently on the "hold" shelf.

The Mill Valley Public Library in California uses McBee Keysort transaction cards and pre-dated charge slips for its circulation operation¹⁶. The system is less costly than a mechanized operation but speedier than conventional procedures. Clerical and professional duties are separated, and the library staff is free to do more professional work. The system has the disadvantages of charging systems in general as discussed above.

The Wayne County Public Library in Detroit also uses a manual punched-card system for circulation control.¹⁶ Eight serially-numbered decks of McBee cards are used for loan or transaction cards; each deck has a different colored edge, one color for each week (Figure 13-12). Charge slips made out by the borrower have space for information about the book and the borrower. A pre-dated (by colored edge) and pre-numbered (edge-

¹⁶ Geer, Helen T., *Charging Systems*, Chicago, American Library Association, 1955.

notched) loan card is placed in the book pocket and the card number copied on the charge slip. These are filed in numerical order. When the book is returned the color on the loan card shows if the book is overdue. Returned loan cards are sorted by color and then each colored deck is sorted numerically. The file is then checked against the file of charge slips. Missing numbers represent overdue books, and notices are sent out. This system requires all books to be due on a given day in the week. The advantages and disadvantages of the system are those described previously for charging systems.

The Library at Wisconsin State College in LaCrosse installed a Keysort charging system on a two-year trial basis. The staff has found that the new circulation procedure takes less than half the time used with the old system, and that errors have been greatly reduced. One disadvantage of the system, it is reported, is a "certain cumbersomeness when a student's library record, when withdrawing from school, has to be cleared". It is necessary to check through the entire student part of the classification file. Even so, the average time for checking a withdrawal is ten minutes¹⁷.

Cards used for such systems as described above are small, ranging from 3 x 5 to 3¼ x 6 inches. There is space for the borrower to write in information about the book being taken. Holes, one or two rows, appear on two, three, or all four sides.

Renewals are handled by re-dating the original call cards. The first due date is covered with a correction sticker and the new date is punched in.

Punched-card files can include other than call number and date due files. Holes can be assigned for a faculty file, reserve book file, bindery file, etc. Some libraries do not keep inactive charges on punched cards, to avoid adding infrequently used cards to the file. Some libraries divide the punched-card file into active and inactive charges, or faculty and student loan, as use or convenience dictate. Other libraries use different colored cards instead of separate files. Again, some libraries use punched-card systems only for books going out for home use. Books to be used in reading rooms or study carrels have paper call slips or conventional call cards. Many special forms of cards and special uses of manual punched-card charging systems are discussed by McGaw² and by Stokes¹⁸.

Analyses and Studies

The application of punched-card systems to library routines results in faster and more accurate processing, as discussed above. By making clerical

¹⁷ Hocker, Margaret L., "Punched-Card Charging System for a Small College Library," *College and Research Libraries*, 18, 119-22, 131 (March 1957).

¹⁸ Stokes, Katherine M., "A Librarian Looks at Keysort," *Library Journal*, 72 (June 1947).

tasks easier, punched-card systems enable staff members to do more professional work. One of the benefits to be derived from this additional time is the ability to analyze the library's operations and to study use of the library's holdings. Such analyses and studies should yield revealing statistics to help the library improve its services.

The Montclair Public Library has used the IBM system extensively for determining the types of books which are being used, where most of the borrowers live, which occupational groups are being reached adequately, and such data. These facts help the library to plan book purchases, to consider location of branches, and to study the reading needs of its borrowers. The Milwaukee Public Library staff uses its IBM set-up to correlate book circulation with information about borrowers, to determine what effect age, sex or education have upon reading taste. A shift of emphasis in purchasing policy can then be effected if necessary. Analysis of total circulation by major classification divisions of the Dewey Decimal System also helps in formulating purchasing policy. Discard analysis is aided by listings arranged by Dewey classes, showing where discarding is heaviest. If the listings are further divided into groups according to the age of the materials, it is possible to determine total holdings in each age group.

Bibliographies of the library's holdings can be prepared by sorting punched cards by subject classification. Materials in a category that cuts across conventional department lines are then compiled for a complete list.

Another use for punched cards has been proposed, that of helping to prepare a cost of books index. The work would involve developing a standard of measurement for book prices for various key years according to subject groupings in terms of a selected base period expressed by index numbers. It is proposed to use 1947-1949 book prices as a base period. Such a price index could be used by libraries for planning acquisition policies, for budget justifications, and the like. A Committee on Cost of Library Materials Index of the American Library Association is working out the details of the proposed project. A tentative punched-card design has been suggested (Figure 13-13). The left side would be checked by the person examining a bibliography for eligible items, and the right side would be used for punching the data. One card would be prepared for each eligible book. The use of punched cards would facilitate the central tabulation of data developed at different locations.

Punched cards have also been proposed as a means of facilitating preparation and management of a Union Catalog of Serials to be established at the Library of Congress. The problems of defining the limits of such a catalog and obtaining cooperative effort in preparing it have been studied by the Joint Committee on the Union List of Serials.

The proposed catalog will contain a listing of the titles and volumes of

year	country			reference	year	country	reference		cost	category	binding	question	cost of books index
							iss	item					
philosophy & religion	1	natural sciences	11		5	6	0	0	0	0	0	0	0
	2		12	number pages	1	1	1	1	1	1	1	1	1
fine arts	3	law	13		2	2	2	2	2	2	2	2	2
	4		14	cost	3	3	3	3	3	3	3	3	3
social sciences	5	medicine	15	bd. unbd.	4	4	4	4	4	4	4	4	4
	6		16	binding	5	5	5	5	5	5	5	5	5
literature	7	agriculture	17		6	6	6	6	6	6	6	6	6
	8		18	question	7	7	7	7	7	7	7	7	7
novels	9		19		8	8	8	8	8	8	8	8	8
	10		20		9	9	9	9	9	9	9	9	9

Figure 13-13. Card proposed for use in preparing a cost of books index by a Committee of the American Library Association.

serials held by the principal research libraries in the United States and Canada (about 500,000 titles and over 50,000,000 volumes). As a by-product of the catalog a *Union List of Serials* would be published every twenty-five years and in-between current service with five-year cumulations. In addition special subject fields can be compiled. A general list by country of origin would also be useful, in light of the growth of area studies.

Punched cards would be used during the period of collection, collation, and listing of the great majority of titles. Lists would then be sent out to cooperating libraries for addition of their holdings. The information secured in this fashion would be consolidated and typed in the center portion of IBM cards, leaving the end sections of the cards free for coding subject and country codes, and numeric codes for maintaining the titles in alphabetical sequence. It has been suggested that an IBM electrostatic printer might be used to reproduce duplicate decks of cards, if needed, or to run off a limited number of special lists, and to provide copy for use as the basis for photo-offset printing of the main list. Since the National Union Catalog symbols for holding libraries would be typed in the center portion of the cards and not in the end sections, information about the holdings of a group of libraries in a particular region of the United States would be difficult to obtain. Preparation of regional lists would probably require the use of additional files of punched cards with the symbols for holding libraries coded into them. To conserve space these subsidiary records could be stored on magnetic tape.

As punched-card systems are used in more and more libraries, additional uses will be found for the data that can be thus handled so conveniently. For example, numerous libraries in organizations or institutions which already have a battery of punched-card machines are using these machines,

rent free and without a budget item for the operators' time, for sorts and summaries in one phase or another of their routines. The application of general business methods will help a library to render its services more effectively.

General References

Kilgour, F. G., "New Punched Cards for Circulation Records," *Library Journal*, 64, 131 (Feb. 15, 1939).

Maxfield, D. K., "Library Punched Card Procedures: Past Experience and Future Possibilities," *Library Journal*, 71, 902-5, 911 (June 15, 1946).

Baehne, G. W., ed., *Practical Applications of the Punched Card Method in Colleges and Universities*, New York, Columbia University Press, 1935.

Parker, Ralph H., "Adaptation of Machines to Book Charging," *Library Trends*, pp. 35-41 (July 1957).

Circulation Control for Libraries, New York, International Business Machines, 1953.

Geer, Helen T. "Charging Machines," *Library Trends*, pp. 244-55 (Oct. 1956).

Library Literature, New York, H. W. Wilson Co. Published quarterly; bound annual volumes and three-year cumulated volumes also published. Contain author and subject index to literature on library science and librarianship. See headings "Machines and the Library," "Punched Cards—Library Uses," and "Equipment and Supplies."

Lowe, Ruth K., "Additional Bibliographic Uses for Keysort Punch Cards," *Library Journal*, 76, 196 (Feb. 1, 1951).

Gull, C. D., "Summary of Applications of Punched Cards as They Affect Special Libraries," *Special Libraries*, 38, 208 (Sept. 1947).

Smith, S. W., "——— And a Few Machines," *Library Journal*, 74, 1044 (July 1949).

Rift, Leo R., "An Inexpensive Transaction Number Charging System with Book Records," *College and Research Libraries*, 18, 112-18 (March 1957).

Chapter 14

REVIEW OF APPLICATIONS

BARBARA L. HAKSTEEN

*National Science Foundation
Washington, D. C.*

Introduction

Research in a specific field invariably involves the need for information which is related to that previously found only after a systematic combing of abstract periodicals and other secondary sources. In order to avoid a repetition of this effort and to obtain information quickly, many groups and individuals are turning to punched-card systems. It is the purpose of this chapter to present brief summaries of published articles and reports which illustrate the way in which a surprisingly wide variety of fields have defined and solved their information problems by adapting the basic principles of punched cards to suit their needs. More detailed information may be found by consulting the references cited.

Nuclear Data

IBM equipment has been used to establish an index referring physicists from nuclear properties to the nuclides possessing them. An IBM card is made for each property (half-life, energy of the various emissions, stability, kind of emissions from active isotopes, availability from AEC, presence of natural radioactivity, etc.) and each nuclide occupies the same punching position on each card. Two sets of different colored cards can be used, one to indicate light nuclides and one to indicate heavy nuclides. The system is indefinitely expansible since any number of properties can be included simply by starting a new card. Selection consists of pulling cards for given properties; the matching holes represent desired nuclides. The entire system is inexpensive since the cards are manually operated¹.

A Keysort card system has also been proposed which reverses the procedure and has one card for each nuclide on which are punched all the properties it possesses. This has the advantage of having all the available information on a particular nuclide on one card. As with the IBM equip-

¹ Wachtel, Irma S., "A Punched Card Index for Nuclear Data," *Am. Doc.*, 3, (1), 56-7, (Jan. 1952).

ment, any type of information can be indexed—general statements as well as specific measurements².

A new research tool has been developed by Bonino and Laing, Pittsburgh Glass Co. Research Laboratory, Creighton, Pennsylvania. It is called the Raychronix Punched Card Nuclide Identifier, and uses 5 x 8 inch McBee Keysort cards. This system enables fast nuclide selection or identification by atomic number, mass number, chemical symbol, stability or radioactivity, availability, types of radiation, half-life and energy of radiations. One set of holes on the card carries the alphabet so that the symbols of the elements can be coded. Both symbols for elements with two names, such as Ra²²⁴ and ThX, are used. The atomic number (Z) and the mass number (A) are coded in a series of fields so that the card for a nuclide with a known A or Z can be searched for and selected from the randomly ordered pack of cards. An example of the particular value of this tool is in the field of health physics where it is desirable to identify unknown radioactive materials so that necessary precautionary measures can be taken without delay. Alphabetic filing is not necessary as cards can be quickly removed from any position in the pack^{3, 4}.

A punched-card system codifying some basic characteristics of radioisotopes has been used by the Western Division of Tracerlab, Inc., Richmond, California. The purpose of this card catalog is to facilitate the identification of unknown radioactive isotopes on the basis of half-life and radiation and to eliminate tedious searching of tables. It identifies isotopes also by their modes of formation, percentage of total radiation, decay schemes, and conversion coefficients. By eliminating stable isotopes and those with half-lives of less than five hours, the number of cards was reduced from 1000 to 375. The system uses McBee Keysort cards. The code chosen involves fourteen simple characteristics which are assigned but one hole each. These simple properties are: emits alpha particles, emits negatrons, emits positrons, emits gamma radiation, emits electrons, decays by K capture, decays by isomeric transition, is a fission product, has a radioactive daughter, belongs to the thorium series, belongs to the neptunium series, belongs to the uranium series, belongs to the actinium series and is naturally occurring. Seven quantitative properties are also coded, such as half-life, maximum gamma, positron, negatron and alpha

² Wachtel, Irma S., "Indexing Nuclear Data on Punched Cards. Preliminary Edition," USAEC Technical Information Service, TID-469, (April 26, 1951); Appendix on page 7.

³ Brochure from Radioactive Products, Inc., 443 West Congress, Detroit 26, Mich. "Rapid Nuclide Identification and Selection by 9 Major Classifications."

⁴ Bonino, J. J., and Laing, K. M., "Punched-Card Classification of the Nuclides," *Nucleonics*, 10, (2), 68, (Feb. 1953).

energy, and predominant gamma and negatron energy. Each of these properties requires a group of four holes to describe the approximate half-life and energies of a given isotope. In addition, a group of eight holes is used to designate the atomic number. Only slightly more than one-half of the available holes are now assigned, assuring the continued and expanding usefulness of the cards⁵.

Biological Data

The applications of punched cards are perhaps more numerous and varied in the field of biological data than in any other. This may be due to the fact that there are a larger number of variables in this field, most of which cannot be controlled as are temperature, pressure and concentration in chemistry and engineering.

The Dow Chemical Co. at Midland, Michigan, is using IBM cards to facilitate coordination between the work of the chemist and the biologist. A combination of numerical and alphabetical coding is used to identify the results of 75 different test procedures on 11,000 compounds. At the present time, approximately 200,000 punched cards are required to record the results of these tests. The system described has combined many of the functions of a biological clearinghouse with routine reporting of current biological tests. The use of nontechnical personnel to handle the bulk of the reporting, correlation, and indexing has greatly decreased the amount of time spent by professional research personnel on clerical work. The primary organization is by chemical entity, using the *Chemical Abstracts* names, a chronological serial number, and a structural classification number. Numerical codes are used for the scientific name of each test, the test method, concentration of the chemical tested, and the biological result of the test⁶.

A bibliography of 5,000 references is maintained at the Roscoe B. Jackson Memorial Laboratory, Bar Harbor, Maine, of all papers on specific inbred strains of mice, named genes in mice, or named transplantable tumors. The references have been classified on 5 x 8 inch Keysort cards (Figure 14-1) and are separated into periods of years (1930-34, 1935-39, etc.). They may be sorted directly by subject or strain individually named in the margins of the card. On the other hand, the cards may be sorted indirectly for which it is necessary to consult a key or index and needle a

⁵ Lukens, H. R., Jr., Anderson, E. E. and Beaufait, L. J., Jr., "Punched Card System for Radioisotopes," *Anal. Chem.*, **26**, 651 (April 1954).

⁶ Dunn, E. E. and Lynn, G. E., "Reporting and Indexing Biological Data by IBM Punched Card Methods," presented at the American Chemical Society Meeting, March, 1952.

The diagram shows a punched card layout with the following sections:

- Top Section:** Classification codes (A through O) and a list of fields (A through O).
- Middle Section:** Fields for AUTHOR, TITLE, and REFERENCE.
- Bottom Section:** A list of minor strains (A through O).

The card is divided into three main sections: AUTHOR, TITLE, and REFERENCE. The AUTHOR section is on the left, the TITLE section is in the middle, and the REFERENCE section is on the right. The card is also divided into three main sections: AUTHOR, TITLE, and REFERENCE. The AUTHOR section is on the left, the TITLE section is in the middle, and the REFERENCE section is on the right.

Figure 14-1. Keysort cards used for bibliographic references at the Roscoe B Jackson Memorial Laboratory.

code number to locate references to the branch of a field of interest or a given minor strain or the named transplantable tumor sought⁷.

A punched-card system was started at Lilly Research Laboratories in Indianapolis to separate for comparison purposes antibiotics with specific groups of properties. The file was originally established as a name file to supplement Baron's "Handbook of Antibiotics." A standard 80-column IBM card is used which contains no special printing and on which are punched 15 classifications of physicochemical data, 5 classifications of biological data in vitro, 8 classifications of biological data in vivo and 5 classifications of biological data in vivo-toxicity. Definite information references supporting the groups punched on the IBM cards are contained on master cards which are different colors to enable quick location in the files. The master cards are related to the proper IBM cards through the corresponding file number. The user who wants information on a specific antibiotic goes directly to the master card file which is arranged alphabetically. The coding outline makes use of positive results only and the file is intended as a guide to the literature, not as a substitute for it. The primary emphasis of the file is chemical but a limited basis for biological comparison is also provided⁸.

Dr. Saul M. Bien, Lynbrook, New York, has devised a system for

⁷ Staats, J., "A Classified Bibliography of Inbred Strains of Mice," *Science*, 119, (3087) 295-296 (Feb. 26, 1954).

⁸ Ohrmund, Margaret, "An Antibiotic Literature File for Chemists," presented before the Symposium on Pharmaceutical and Medicinal Literature, Division of Chemical Literature, American Chemical Society, Sept. 16, 1954.

unpunched: present
punched: missing
deep punched: extracted

1: normal
2: local irritants
3: hypertrophic

4: recessive
5: ulcerative
6: periodontitis
7: periodontosis

1: leptoprosopic
2: mesoprosopic
3: euryprosopic
4: gross asymmetry

1: dolichoccephalic
2: mesiocephalic
3: brachycephalic
4: gross asymmetry

1: 23° 6: 54° 11: 66°
2: 42° 7: 57° 12: 68°
3: 45° 8: 60° 13: 70°
4: 48° 9: 62° 14: 72°
5: 51° 10: 64°

1: 50 6: 100 11: 150
2: 60 7: 110 12: 160
3: 70 8: 120 13: 170
4: 80 9: 130 14: 180
5: 90 10: 140 pounds

SAUL M. BIEN, D.D.S.
654 MADISON AVENUE
NEW YORK 21, N. Y.

unpunched: tooth in normal relationship
punched: tooth grossly out of position in the arch
deep punched: tooth impacted

Figure 14-2. Keysort cards used for registering orthodontic diagnostic data by Dr. Saul M. Bien.

registering orthodontic diagnostic data on hand-punched and hand-sorted McBee Keysort cards. (Figure 14-2). A total of 305 different items abstracted from the patient's history, physical examination, radiographs, photographs and models may be punched on the card in code. The teeth are numbered according to the universal system, starting with the upper right as number one and ending with the lower right as number 32. A punched position on the cards indicates the absence of the tooth assigned that number. Cephalometric and other anthropometric data can be recorded on the face or back of the card⁹.

Bucknell University has established a file of edge-notched cards which constitutes a bibliography of references to studies published on the golden hamster. The cards have been prepared so that authors may be arranged in alphabetical order, or cards may be selected according to journal, date of publication, author and subject. References on bacterial diseases, cancer,

⁹ Bien, Saul M., "Registration of Orthodontic Diagnostic Records for Statistical Evaluation," *Am. J. Orthodontics*, 41, (6), 482-483 (June 1955).

caries, parasitology, pathology, and virus diseases account for most of the total publications. The file has proved very useful in the investigation of the genetics of the hamster, and in pointing out that very little is known about the requirements for optimal growth and reproduction, or anatomy of the hamster¹⁰.

A description of a coding system using 3 x 5 inch punched cards has been reported by Norman D. Levine, College of Veterinary Medicine and Agricultural Experiment Station, University of Illinois. This system has been used to set up a file of several thousand cards for handling abstracts of veterinary, medical and general parasitology using issues of *Biological Abstracts* as a basic source. A decimal code covering three fields of four holes each is used to code the most important parasite genus discussed in the paper. This permits the use of 999 numbers in the code. A second set of three fields is used for a second parasite genus. A decimal code covering two fields and allowing for 99 numbers is used for the last genus. The first and second subjects are assigned five holes each and cover such items as cultivation, diagnosis, evolution, excretion, genetics, growth, regeneration, etc. An additive code based on five holes is used for the first letter of the first author's name. Finally, three single holes are punched separately when the paper discusses more than two parasite genera, more than one host, or more than two subjects, respectively. This leaves a group of seven holes which are available for other information one might wish to code¹¹.

A systematic review of the world literature on vision in invertebrate animals was begun in 1946 by Lorus and Margery Milne, University of New Hampshire, on hand-notched and hand-sorted 3¼ x 7½ inch Keysort cards. At present, abstracts have been typed on more than 4,500 of these cards which have been coded according to a system devised by the authors. The card file is arranged so that it can be sorted alphabetically by author and then chronologically under each author's name, by subject and chronologically under each subject, and by taxonomic group. Twenty holes are assigned to code the first four letters of the senior author's name. Additional cards without abstracts are prepared for junior authors with a cross reference to the senior author's card. Nine holes are used to code the date of the reference; two fields of 7, 4, 2, 1 to cover the units and decades, and a ninth hole to be notched for any date earlier than 1900. Subject categories were punched in sixteen holes and coded according to the "Classification of Zoological Literature" used by the Wistar Institute Bibliographic Service. A single hole is punched to indicate that an abstract has been made

¹⁰ Magalhaes, Hulda, "The Golden Hamster as a Laboratory Animal," *J. Animal Technicians Association*, 5, (2), 39-44 (September 1954).

¹¹ Levine, Norman D., "A Punched Card System for Filing Parasitological Bibliography Cards," *J. of Parasitol.*, 41, (4) 343-352, (August 1955).

and the reference checked for accuracy. Thirty-three holes remain for taxonomic use, twenty of which are used for the first four letters of the generic name. Dewey's system of decimal classification was followed to code for taxonomic position, the numbers being abbreviated by deleting the initial (common) 59. Thus the Dewey 593.1 for Protozoa became 3.1. Some material was typed on the face of the card: authorship, title in the original language, journal reference as it appeared on the title page of the volume rather than in the form used in the *Union List of Serials*, the abstract and the source of the reference. The code for the source of the information was a simple one. For example, 24 Brown 627 would indicate that the reference was cited on page 627 of a paper written by Brown in 1924.

From the standpoint of cumulative experience with this system, the authors have several suggestions on ways in which it could be improved (Figures 14-3 and 14-4). For example, E-Z Sort cards with six holes per inch could be used instead of the Keysort which have only four per inch. Thus the same size card would accommodate 50 per cent more holes and

Notched single depth - microfilm of part, double depth - of whole
Notched single depth - seen, double depth - copy in file
Notched single depth - 1700's, double depth - 1800's, double depth - 1600's

Spares

First Second Third X100 Decade Year Initial Second Third Fourth

Journal Abbreviation Date of Publication Author's Surname

Sigmata
Photoreceptors: not found/neuronal
Eyespots: unicellular/compound
Ocelli: simple/compound
Stemmata/compound eyes
Camera-style eyes
Accommodation
Adaptation (light and/or dark)
Latency
Intensity discrimination
Visibility acuity
Resolution acuity
Polarization-plane detection
Motion acuity & flicker fusion
Binocular field
Courtship and vision
Pollination and vision
Form & pattern perception
Obliterative coloration
Photokinesis
Phototaxis
Spectral sensitivity: 390-760 mμ
Spectral sensitivity: above 760 mμ
Spectral sensitivity: below 390 mμ
Photodynamic action
Photoperiodism
Pigment migration: in eye/in body
Tapetum
Electroretinograms/nervous impulses
Threshold

Protista
Sarcodina
Ciliata
Hydrozoa
Scyphozoa
Anthozoa
Tentaculata
Trematoda
NEMERTINEA
Rotifera
Gastrotrocha
Kinorhyncha
Nematoda
BRYOZOA
Amphineura
Gastropoda
Pelecypoda
Cephalopoda
Archimedeida
Polychaeta
Oligochaeta
Hirudinea
CHAETOGNATHA
TARDIGRADA
ONYCHOPHORA
Brachiopoda
Crustacea
Copepoda
Brachyura
Cirripedia
Malacostraca
Trilobita
Xiphosura
Euryptera
Arachnida

Chordata
Mammalia
Aves
Reptilia
Amphibia
Teleostei
Chondrostei
Agatha
Cephalochordata
Urochordata
HEMICHORDATA
Ophiuroidea
Echinodermata
Echinoidea
Asteroidea
Holothuroidea
Crinoidea
Holocephala
Hemimetabola
Ametabola
Diptera
Chilopoda
Pyrogastridae
Glyptidae
Araneidae
Acari
Phalangida
Pseudoscorpionida
Scorpiones

Spares

Figure 14-3. Keysort card used for controlling literature on vision in invertebrate animals.

Matched for any date not in 20th century

Date of publication

Decade Year Initial Second Third Fourth

Author's surname

Initial Second Third Fourth

Matched for 2nd half of alphabet in

Photobiology

Spectral sensitivity

Optics

Anatomy

Physiology

Neurophysiology

Nerve impulses

Orientation

Kinetic responses

Photoperiodism

Pigment migration

SMITH, H. G.

1938 Apr. 27

The receptive mechanism of background response in the chromatic behavior of crustacea.

Proc. Roy. Soc. Lond., Ser. B, 125 (839) : 250-263, 9 figs., pl. 12

55 McWhinnie & Sweeney 174

Ligia oceanica

Tried effects of limiting the visual field; found one group of ocelli (OVER)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Generic name

Classification code

Matched when article read and abstracted

Figure 14-4. Keysort card used for controlling literature on vision in invertebrate animals.

the percentage assigned to author (now 25 per cent) would fall sharply. If a card with two rows of holes were used it would mean a further saving in space, but any name having the same letter in two or more of the first four positions could introduce confusion. Another E-Z Sort card with four rows of holes could be used, notching the initial letter of the author's name single depth, the second letter double depth, and so on. Different systems have been suggested for expanding the alphabet to 30 characters or simplifying it to 23 characters. Another improvement discussed was the possible use of ten holes for the date in order to indicate the century more definitely. This would mean an increase of two holes, the first of which would represent the 20th century if uncut and the 19th if cut, while the second hole would indicate the 18th century if cut and the 17th if uncut. An additional twelve holes would improve the manner of coding the journal reference using a three-letter designation with double-row perforations. For example, JAB would stand for the *Journal of Animal Behavior*, JGP for the *Journal of General Physiology*, etc. Another hole to indicate that a copy of the reference is in the file and one to show that a microfilm copy is in the file would also add to the usefulness of the file. The Dewey Decimal system proved to be of little value in this applica-

tion. Also the use of twenty holes for the generic name was wasteful of space. The number of papers discussing a variety of organisms grew beyond expectations. For these, only direct coding would have been helpful. It is estimated that a list of 55 phyla, subphyla, classes and subclasses would handle all of the great variety of animals upon which photosensory studies have been published¹².

IBM punched-card methods have been perfected by the Chemical-Biological Coordination Center, at the National Research Council in Washington, for coding information concerning the biochemical transformations undergone by pesticides in the course of their metabolism. A code based upon a comprehensive classification of enzymes permits the recording of the effects of pesticides upon these catalysts. This information can be retrieved by searching for the type of reaction, the organ, the species, the names of the pesticides, and their products. An abstract suitable for coding is prepared on a code sheet form. The information is broken down into coding fields, such as taxonomy, organ, host, specific effect, general effect and dose level. Punched cards are then prepared¹³.

Photography

A bibliography on photographic theory originally prepared at Eastman Kodak has been established on McBee Keysort punched cards. The bibliography covers emulsion making, latent image formation and development. Color photography and sensitometry are not included, except to a minor extent. The card used measures $6\frac{1}{2} \times 7\frac{1}{2}$ inches and is specially printed to indicate holes for coding the senior author's name, the date and type of publication and the information mentioned above¹⁴.

Keysort punched cards have also been used by the Edwal Laboratories, Inc., Ringwood, Illinois, for a file of photographic references in which abstracts are pasted to the cards. The file may be searched by main subject, author's name, date of publication or patents. Main subject coding is numerical according to the Universal Decimal System as used by both *Kodak Abstracts* and *Photographic Abstracts*, the two abstract journals

¹² Milne, Lorus J., and Margery, "Foresight and Hindsight on a Punch-Card Bibliography," report on work done at the University of New Hampshire, partially supported by the graduate school and at the Scripps Institution of Oceanography, University of California. Contribution from the Scripps Institution of Oceanography, New Series, No. 968.

¹³ Wood, G. C., and Welt, I. D., "A Multi-indexed Machine Sorted, Punch Card System for Pesticide Metabolism Data," *Agriculture and Food Chemistry*, 4 (10), 886-888 (Oct. 1956).

¹⁴ LuValle, James E. Item #22 in "Abstracts of Presentations. Investigators Restricted Seminar #1 on the Chemistry of Photographic Processes," Chicago, Illinois (Sept. 4, 1953) Sponsored by the Chemical Division of the Headquarters Air Research and Development Command.

used as the main source for the reference file. The alphabetical coding of the author names is according to the revised method of Cox, Bailey and Casey, *Chemical and Engineering News*, September 25, 1945. Another suggested use for these cards is as an index to Kodachrome medical slides. These can be filed according to pathological condition, part of the body affected, symptoms, etc.¹⁵

Punched cards have also come into use in the filing of photographic negatives. E-Z file cards with Filmsort positive transparencies of the print have been used by Eastman Kodak for this purpose. Different classification systems are necessary to suit the type of work being done by individual photographic organizations. For the commercial photographer, the name of the subject or client serves as the best filing key. This is true also for such files used for police identification, and by industrial, personnel passport and medical photographers. Commercial illustrators and industrial photographers, on the other hand, are more concerned with jobs or products and for them filing by product name, company or department, job name or job number is more desirable. Coded punching around the edge of the card carries the key to the filing classification for the photograph covered by each card and allows rapid mechanical selection of the desired cards. Print and data sheet transparencies may be examined directly from the card with a viewer without removing the negative from the file. This system fulfills the two most important requisites for filing negatives: prompt location and minimum handling to avoid scratching the surface and damage from dust particles¹⁶.

Laboratory Records

One of the most important factors in the successful operation of an industrial analytical laboratory is an adequate system of keeping records of samples being analyzed and of the data obtained. A system described by A. H. Hale and J. W. Stillman of the E. I. du Pont de Nemours & Co., Inc., in Wilmington, has been developed to meet this need and is in actual use in several industrial laboratories. Printed slips with interleaved carbons are used for recording and reporting analytical data. From these slips punched cards are prepared, first to provide positive control of the progress of the analysis and then to locate the original data on the report slips in the permanent file. McBee Keysort cards (5 x 8 inch) are punched to show the assignment of samples to be analyzed, the progress of the analyses, the location of the analytical report in the files and to provide information on the operation and efficiency of the laboratory. Code num-

¹⁵ Hill, Thomas T., "Finding Photographic Information," *J. Biol. Photographic Association*, 17, (3) 103-114, (March 1949).

¹⁶ "Filing Negatives and Transparencies", a twenty-page pamphlet prepared by Eastman Kodak Company, Rochester, New York, (Oct. 1953).

bers are used to designate each chemist so that it is possible to determine the name of the worker from the number punched in the upper right corner of the card. The year and the month of receipt of the sample are punched in the upper left corner. The material block has space for 4,000 numbers to include all kinds of samples received. Five-hundred numbers are assigned to organic compounds and 700 to inorganic compounds. There are sections to notch for polymers and their modifiers (colors, fillers, inhibitors, etc.)¹⁷.

Marginally-punched cards are used by the Strong Cobb Company of Cleveland, Ohio, as part of the record-keeping system for its pharmaceutical control laboratories. A 6 x 8 inch Keysort card serves as the permanent record of all analytical control data for each product manufactured. The card is custom printed to specification thus eliminating the unnecessary copying of information which was formerly transcribed by hand for each job. Space is provided on the printed portion of the card for the date of receipt of the sample, the stage of production, assay data, etc. Information is coded for all constituents assayed, manufacturing difficulties encountered, and the type of product being developed. A similar system is used by the White Laboratories of New Jersey^{18, 19}.

As long as ten years ago, workers at Eastman Kodak Co. in Rochester realized that some sort of a punched-card system was essential to locate more quickly the approximately 1,000 organic compounds that had been synthesized in the laboratory for research purposes. It was preferable to locate classes of compounds rather than individual substances. Different colored 5 x 8 inch Keysort cards (Figure 14-5) were printed to certain specifications; each color represents one broad group of organic compounds. These cards are filed separately according to colors. Structural groups and features are listed in two rows along the top of the card, together with a few types of substances of special interest such as ureas. Along the right-hand edge are numbers referring to ranges of light absorption, while those on the left refer to numbers and positions of substituents in simple cases. Halogens and amines (primary, secondary and tertiary) are also included on this side. The bottom is reserved for ring sizes, some groups of substances, a few less frequently encountered structural features and the very important section, labeled "hetero." No provision has been made for salts (anions) since this was unimportant in the present application. The order of precedence of groups (and so, of card colors) was arbitrarily

¹⁷ Hale, A. H., and Stillman, J. W., "Development of an Efficient Analytical Record System," *Anal. Chem.*, **24** (1) 143-149 (Jan. 1952).

¹⁸ Naimark, G. M., and Prindle, R. F., "Pharmaceutical Control Laboratory Record System," *Anal. Chem.*, **26** (4) 645-647 (Apr. 1954)

¹⁹ Naimark, G. M. "Industrial Analytical Record Keeping," *Drug and Cosmetic Industry* (Sept. 1955).

Figure 14-5. Keysort card used to record information on organic chemical compounds at Eastman Kodak Co.

Technical Services

²⁰ Allen, C. F. H., "Keysort Punch Card System As Used in the Organic and Polymer Chemistry Department, Chemistry Division, Eastman Kodak Company." April 25, 1956. Report submitted to J. W. Perry by the author, May 11, 1956.

ordinary laboratory report. Fifteen years of records are now coded into 5,000 cards. Within 30 minutes reference to all of the work previously done in a specified area can be located²¹.

Micro Switch, Freeport, Illinois, has initiated a project to determine how product histories, or abstracts of company "know-how," should be prepared and what would be their long-range value. One of the major tasks in the preparation of these histories is accumulation of the information needed. A ready means of retrieving all the available data on such products is essential. As bits of knowledge come in from such divisions as Products Research, Engineering, Methods, and Quality Control, they are recorded on specially designed punched cards (Figure 14-6). The proper classification, coding and punching of the cards make it a relatively simple matter to sort out all information on a given product, on any given process, on quality control data and types of materials. This eliminates duplication of files and cross indexing²². The author's name code is based on a 100-division alphabet breakdown and is punched in two fields on the left side of the card. At the upper right two fields are devoted to coding the product group. One field gives the type of switch involved and the other the catalog listing. The various parts of the switches such as springs, anchors, or plungers are punched into two fields at the bottom. In the two fields next to this, are coded all processes (such as welding or heat treating). Another two fields at the bottom are designated for coding materials involved. Specifications (resistance, ductility, etc.) and a miscellaneous classification use two fields each on the right side. All items in each classification are numbered consecutively from one. The punched card carries the identifying file number for locating the original information.

Market Research

An edge-punched card system was adopted by Magnaflux Corporation of Chicago as a logical approach to the task of storing and retrieving details on current records of customers and prospective customers. Specially designed 5 x 8 inch E-Z Sort cards with 140 holes are used (Figure 14-7). Both sides of the card are printed; white cards are used for customer information and buff cards for prospects, and both of these are filed separately. There are approximately 1,200 customer cards—300 active and 900 inactive. The field in the upper right-hand corner is used to code the first two letters of the company name. Letters A through Mc are assigned numbers one through thirteen. Letters M through Z are given numbers V, V1, V2 etc. through V13. A four-digit number is coded in standard

²¹ "Punch Cards Up Tech Service Productivity," *Chemical Week*, (Nov. 3, 1956).

²² Rhynders, R. W., "Product History Clears Haze from Technical Records," *Industrial Laboratories*, (Feb. 1956).

NAME CODE												YEAR												MONTH												PRODUCT GROUP											
GENERAL FILE NO. EM - 4357												DATE July 25, 1957																																			
SPECIAL FILE NO.												NAME J. Hart																																			
CATALOG NO. 1E046												ADDRESS Design Engineering																																			
QUANTITY																																															
New Design of Plunger (EO-16012) Refer to Engineering Drawings EO-11049 and EO-16012. Larger Collar on Plunger prevents shaft from vibrating out of alignment. New Welding technique superior to soldering because of more positive seal, no flux contamination, and better control of heat conduction in switch. Seal ring material change from S-2 Silicone to S-1416 compound means less wear on the seal.																																															
MICRO SWITCH																																															

NAME CODE												YEAR												MONTH												PRODUCT GROUP											
GENERAL FILE NO. EDM-MP-3061												DATE Aug. 15, 1957																																			
SPECIAL FILE NO.												NAME W. Millard																																			
CATALOG NO. 12E22												ADDRESS Process Engineering																																			
QUANTITY																																															
Potting Terminal - Lead Assembly of Subject Switch. Experimental Trials of the following compounds: 1. - X-421-T - Poor Adherence 2. - X-474-T - Good Bond 3. - X-501-T - Slow Set Vibration and shock test are in process, early indications look favorable for compounds 2 and 3.																																															
MICRO SWITCH																																															

Figure 14-6. Keysort cards used for recording information from company reports at Micro Switch.

STANDARD INDUSTRIAL CLASSIFICATION										CUSTOMER (ALPHA)									
Detroit CUSTOMER NAME: Master Casting Co. ADDRESS: 8260 N. Tenth St. CITY: Benton Harbor STATE: Michigan TELEPHONE: 454-9216 PERSONNEL: E. Equist, Chief Insp. J. Bonner, Recv. Insp. P. G. Chandler, Ch. Met.										SPECIAL PRODUCTS METHOD: Magnatest EQUIPMENT: FM-100 HOW USED: Measure Conductivity of Transformer Secondaries DATE PURCHASED: 4/28/54 <i>shpl 4/15/54</i> LEADS: MX OTHER METHOD: MX ADDRESS BY CORP.: CHICAGO FIELD MARKET ANALYSIS: CHICAGO									
SALES CALLS: 11/5/53 2/15/54 CHECK OF CALLS: 6/22/54 8/24/55 SERVICE CALLS: 5/11/55 COMMENTS: 6/22/54 Installed Unit. 5/11/55 Reported difficulty in zero balance. GBE changed R-21 by adding 250K resistor. No further trouble. 8/24/55 Unit operating ok																			

Figure 14-7. Card for recording information on customers at Magnaflux Corporation.

six-hole fields at the top center of the card to indicate the name of a company in a given area of business. The business or individual is identified in a numerical method by the product manufactured or service rendered using the Standard Industrial Classification. The field on the right-hand side of the card is devoted to the inspection methods marketed by the corporation. Specific equipment is represented in the next five fields on

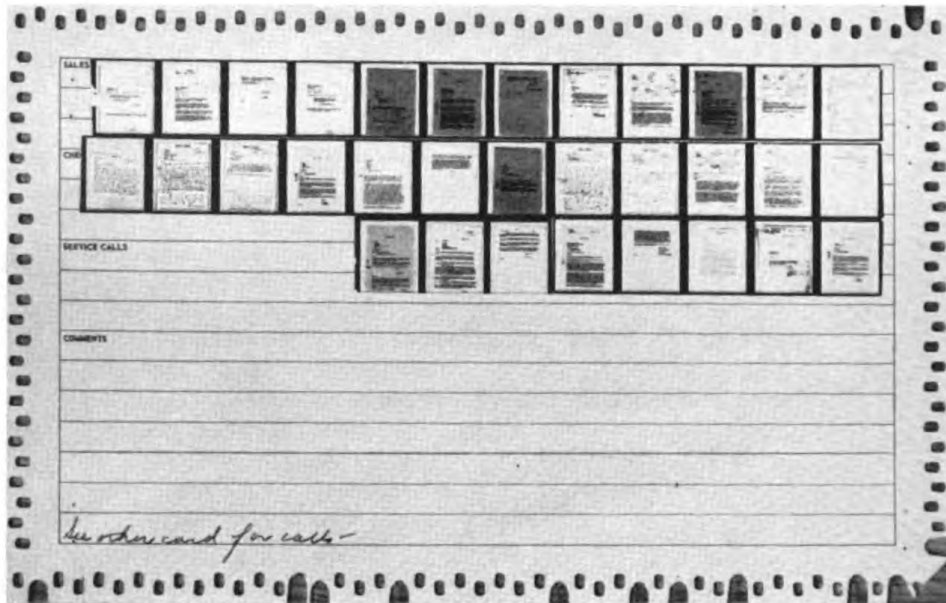


Figure 14-8. Card containing microfilm copy of customer correspondence, as used at Magnaflux Corporation.

the bottom and lower left-hand corner of the card. Each correspondence file is transferred to the individual customer card by microfilming the documents on adhesive-backed film (Figure 14-8). In this way all of the documents pertaining to the customer's file can be scanned by turning over the punched card²³.

Meteorological Data

IBM punched cards have been used in handling large masses of experimental data resulting from upper atmosphere research. Hand-sorted cards could have been used but their capacity is comparatively limited. IBM bibliography cards may be selected by author's name, publishing agency or journal name, date of publication, security classification, language or subject matter. Abstracts of the articles are typed on the back of the card. A recommendation by Boston University's research groups was made for the establishment of a central agency to abstract, index and distribute the cards²⁴.

In 1948 the British Meteorological Office began to index upper air data using Hollerith card systems. Observations were punched on the

²³ Cannon, W. A., Jr., "A Punched Card System for Technical Liaison, Sales Analysis, and File Reductions," Hathaway Instrument Division, Hamilton Watch Co., Chicago, Illinois.

²⁴ Low, Ward C., "Technical Publication Abstracts on IBM Punched Cards I," Technical Note #15, July 14, 1952. Upper Atmosphere Research Laboratory, Boston University.

cards at the observation stations and then sent to the central office for machine sorting and tabulation. It was anticipated that a little more than 130,000 cards would be prepared each year to make available all statistical information about upper air conditions over the British Isles. Coded information includes date, time, place, pressure levels, wind direction, temperature and humidity²⁵.

Hollerith cards were adapted in 1950 to evolve linear-function tables for upper air data. Owing to the large range of values and the varying number of observations, 40,000 totals had to be given to cover the ranges 100° to -100°F and 31 to 11 observations. It is doubtful that the preparation of such tables, as invaluable as they are, would have been worth while by clerical workers as too much time would be lost in the mechanics of working out monthly mean values²⁶.

In 1951 Hollerith cards were used to great advantage in the field of marine meteorology and they made possible the preparation and publication of climatological atlases of the oceans. After various meteorological elements included in the observations were punched in appropriate codes, the cards were sorted and filed in packs according to the month and the 10° Marsden Square in which they belong. A total of approximately 3½ million British and 6½ million German cards were filed in the Marine Branch at that time. The effect of wind velocity on the sea and air temperature, the diurnal variation of the sea and air temperatures in relation to cloud amount and to the sea and air temperature differences, and the effect of wind velocity on relative humidity were investigated^{27, 28}.

Geological Data

Geologists, like other scientists, have been losing ground in their efforts to keep informed on latest developments in their realm of specialized knowledge. In order to help alleviate this situation the Petroleum Research Corporation of Denver, Colorado, has developed and is producing the Micro-Research-Card Library of the Rocky Mountain region. This reference collection is available for purchase or rental. The library contains microphotographic reproductions of approximately 8000 published and unpublished articles and theses. Included are papers from more than 80 periodicals, publications and unpublished reports of the U. S. Geological

²⁵ DeWar, D. "The Hollerith Card System Applied to Upper Air Data," *Meteorol. Mag.*, 78, 163-166, (1949).

²⁶ DeWar, D. "Preparation of Linear-Function Tables on a Hollerith Tabulating Machine," *Meteorol. Mag.*, 79, 137-140 (1950).

²⁷ Gordon, A. H., "Development of Modern Techniques in Marine Meteorology," *Meteorol. Mag.*, 80, 78-83, (1951).

²⁸ Gordon, A. H. "Adaptation of Mechanical Sorting and Tabulating Machines to Research in Marine Meteorology," *Meteorol. Mag.*, 80, 269-270, (1951).

Survey and an almost untapped wealth of theses. The material in this file is so arranged that it can be sorted, in a matter of a minute or so, by geologic subject, by area, or by geologic time. Any given article can also be selected by author. The file now contains about 8000 5 x 8 inch film transparencies, perforated at one end with 207 small holes. Each hole represents a geologic subject, a geologic time, date of writing, or geographic area within the Rocky Mountains. Every article has been coded by a geologist whose notations appear on the card in microfilm form and also as slots extended from the holes in the punched end of the card. Cards are needle sorted. The number of categories needed for optimum separation was determined by the subjects and areas covered. Each card is numbered and filed according to the area its subject encompasses. Numbered areas within the greater Rocky Mountain region are indicated on a map in the library. In addition to the numbered breakdown by areas, letter designations (N, S, E, W, and C, for north, south, east, west and central) indicate the portion of a numbered area involved. To facilitate selection by area further, a separate coding is made by states. Eighty-seven geologic subjects are coded, including structural contour maps, paleontology, radioactive minerals, origin, oil analysis, etc. If a given article is sought, a printed bibliography lists, by author, all articles included in the library and gives the "call number" which is slotted on each card. The first two digits of the call number are the area designations and these are slotted along one edge of the card so that a card out of position in the file is detected at once. Based on the demand, as indicated by the first effort, the library may be expanded to coverage of the continent, and even of the world²⁹.

The accumulation of data resulting from a program undertaken by the Ohio Division of Geological Survey to evaluate the coal reserves of the state bed by bed has been compiled on punched cards. The system of tabulation selected utilized IBM equipment (a key punch machine for punching data into the cards, a sorter which mechanically sorts the cards into any desired order, and a tabulating machine which mechanically prints the data from the punched cards). The file contains approximately 12,000 individual outcrop records from the 25 coal-bearing counties of the state. The tabulation procedure was designed first to study the problems in the Pennsylvania part of the Ohio geologic section. The transfer of the information from the original file sources to punched cards began in 1954. For the 14 counties evaluated so far, over 4,000 stratigraphic records have been coded containing more than 10,000 observations of individual coal beds and their associated strata. The code as set up uses a four-digit stratigraphic number for each coal bed and a one-digit lithologic number within a single coal-to-coal interval. The lithologic number consists of a

²⁹ Chronic, John, "How Microfilm Library Aids Research," *World Oil*, (May 1956).

one-digit number for each of the nine more common rock types such as marine limestone, sandstone, etc., which commonly occur in the interval between coal beds. The four-digit stratigraphic code serves to designate geologic age and position within the standard geologic column of Ohio. The one-digit number code serves to designate the lithology and position of the strata within the depositional cycle comprising the interval from the base of one coal bed upward to the base of the next younger coal bed. Counties are coded numerically from 01 to 88 in accordance with their alphabetic position and townships are coded numerically in accordance with their alphabetic position within the county. Every digit of the standard 80-column IBM card is used in the general part of the study, necessitating the use of a second card for tabulation of interval data. Identification data such as file number, sources and location are repeated for each cycle, so that reference to original sources may be made at any point regardless of the subsequent classifications in which any one cycle appears. A second card is punched for each interval in the section. A competent operator can assemble data at the rate of 700 to 1000 cards per day.³⁰

Astronomical Data

Lick Observatory at the University of California is engaged in the preparation of a list of all double star measures beginning with the year 1927.0, the closing date of Aitken's "New General Catalog of Double Stars." All measures for double stars of the Northern Hemisphere that have appeared in print or in manuscripts have been entered on more than 80,000 IBM punched cards. The cards are proving so useful and efficient for compiling lists and carrying out statistical studies that the measures cataloged for many years at the Southern Hemisphere are being sent to the Observatory from the Union Observatory in South Africa, to be punched. Other files maintained at Lick Observatory include punched cards for some eclipsing binaries. New projects for computing are being planned which by former methods would take more than an astronomer's lifetime. Other applications in the field of astronomy include the extended moon and planetary ephemerides that have been constructed by Brouwer, Eckert and Clemence, the automatic plate measuring machine developed at the Watson Scientific Computing Bureau by Eckert and his associates, and the use of punched-card processes to calculate the orbits and ephemerides of comets and minor planets by Herget and Cunningham³¹.

³⁰ Smith, William A.; Brant, Russell A.; and Klein, Marian S., "An Application of Business Machine Technique to Stratigraphic and Coal Resources Studies," Information Circular No. 18, State of Ohio Dept. of Nat. Resources, Div. of Geological Survey, 1956.

³¹ Personal Communications: Feb. 24, 1956, Mrs. James F. Chappell, Lick Observatory to Robert S. Casey and March 23, 1956, H. M. Jeffers, Lick Observatory to Robert S. Casey.

Legal Data

A special committee of the New Jersey State Bar Association made an investigation into the realm of mechanized and automatic literature searching, an explanation and resumé of which was published in 1953³². This investigation showed that a crisis existed in the storage and use of legal literature to a degree which demanded an immediate solution whereby the material read by human eyes and brains could be examined and selected at a rate at least 100 times greater than that of which humans are capable. In his article, Biunno suggests a guide for preliminary experimentation and discusses the use of a machine language called "Luko" in which 20 consonants (Q being excluded) and 5 vowels are arranged in all possible combinations to provide code headings.

The American Bar Foundation of Chicago, Illinois, conducted an experiment designed to demonstrate one possible adaptation of punched cards and mechanical retrieval to legal research processes as performed daily throughout the country by lawyers and judges. Remington Rand cards and equipment were used to demonstrate this system at the Symposium on Systems for Information Retrieval, held in Cleveland in April 1957. The Illinois Divorce Statute was used as the master code and the cards were punched on the theory that the Illinois lawyer would want to know only to what extent the Idaho law, for example, is different from the Illinois law with which he is most familiar. A sheet was prepared showing the correlation between the holes in their numbered positions on the cards and the master code. The American Bar Foundation also hopes to use punched cards and the Remington Rand installation in the membership department of the American Bar Association to index the publications of professional legal organizations which are not now covered in the Index to Legal Periodicals³³.

A suggested coding method for legal data using a specific area as a test case (liability of electric power and telephone companies for injury or damage by lightning transmitted on wires) has been received from Charles Cobb, Jr.³⁴. The plan would use 5 x 8 inch E-Z Sort cards with two rows of holes along each edge (Figure 14-9). An analysis suitable for all purposes would require the use of a logical scheme adapted to the expression of the

³² Biunno, Vincent P., "Searching Legal Literature—An Appraisal of New Methods," *Law Library Journal*, 46, (2), 110-119, (May 1953).

³³ MacKinnon, F. B.; Leary, J. C.; and Levinson, D., Jr., "An Analysis of the Problem and An Experimental Adaptation of Punched Cards and Mechanical Retrieval to Legal Research and Indexing of State Statutes, Codes and Session Laws," Prepared for the Symposium on Systems for Information Retrieval, April 15-16, 1957, Cleveland, Ohio.

³⁴ Personal Communication, June 20, 1956 from Charles K. Cobb, Jr., Law Book Department of Little, Brown and Company, Boston, Mass. to James W. Perry.

The card is a standard 80-column punched card. The fields are as follows:

- Type**: 1-4 (10 columns)
- Degree**: 1-4 (10 columns)
- Cause**: 1-4 (10 columns)
- Defendant (II)**: 1-4 (10 columns)
- Plaintiff (III)**: 1-4 (10 columns)
- Means (IV)**: 1-4 (10 columns)
- Contributory negligence (V)**: 1-4 (10 columns)
- Date**: 1-4 (10 columns)
- Negligence (IA)**: 1-4 (10 columns)

Figure 14-9. Card used for recording legal data by Charles Cobb, Jr.

intentions of legal propositions. Four numerical fields would be used to describe the defendant (in this case one number for electric power companies and another for telephone companies), the plaintiff (a customer or member of the general public), the means by which the injury or damage was inflicted, and the nature of the injury or damage. The cases gathered from the legal digests and other sources are also sorted by state and are arranged chronologically for each state. The system would be adequate for sorting cases by legal result, both for final judgment and for the conditions of liability.

Chemical Literature

The great bulk of existing chemical literature and the continued high rate of production have greatly accelerated the search for some method or device to permit a rapid and thorough search of the whole breadth of this literature to verify the presence or absence of a given fact. A suggested system has been described in the literature³⁵ which combines a microcard with a punched-card coding and sorting system. A standard 3 x 5 inch microcard is fitted into an area covering 60 columns of an 80-column IBM punched card. The remaining 20 columns are available for coding. Ten columns are punched with a standard Dewey-Decimal or other similar classification of the main subject matter or title of the paper reproduced on the microcard. The remaining ten columns are punched with a set of

³⁵ Williams, T. J. and Rose, A., "A Solution to the Problem of Storage and Availability of Chemical Literature," *J. Chem. Educ.*, **29**, 146-147, (March 1952).

according to source which permits the assignment of a number to each source as it appears. The holes labeled B through H at the top of the card are reserved for the source code, allowing a total of over 2,200 sources to be coded if necessary. In the date code the first outer hole I_0 corresponds to 1800 and the first inner hole, I_1 , to 1900. The remaining holes in the group J through N, provide spaces for the ternary code used for the author and source. The subject code numbers are assigned to the outer and inner row of holes on the bottom and left-hand side of the card, which are numbered from 1 to 46. The major division referring to specific plating practice is located from numbers 31 to 36, outer and inner holes. The subtopics are acidity, addition agents, agitation, temperature, etc. The code numbers for the metals are assigned to the outer and inner rows of holes along the right-hand edge of the card labeled with the chemical symbols for some of the elements. The reference and an abstract are typed on the face of the card.

Stanley Kirschner, Department of Chemistry, Wayne State University, Michigan, has described his system for coding and abstracting chemical literature. He uses a specially designed IBM card (Figure 14-11) which allows much of the important information to be coded directly by punched holes and provides space for the title and a brief abstract on the face of the card. Two initials and the senior author's surname may be punched directly into columns 1-10 on the card using the IBM letter code, which allows coding any letter of the alphabet by means of a double punch in a single vertical column of numbers. The name of the journal in which the original reference appears is coded into columns 11-14 by means of a four letter abbreviation such as that devised by Bishop for the "Codex" system [C. Bishop, *Am. Documentation*, 4, 54 (1953)]. The volume number, page and year may be coded directly into columns 15-24 using the ap-

Figure 14-11. Punched card used for recording codes pertaining to chemical literature by Stanley Kirschner, Wayne State University.

propriate numerals. Only the last two digits of the year are coded. The name of the journal in which the abstract appears may be coded into columns 25–28 in a manner analagous to coding the name of the journal in which the original reference appeared. The volume number is coded directly into columns 29–31. Columns 32–36 are used to indicate the column or page number for the abstract and column 37 may be used to show the location of the abstract on the page. The individual subjects are coded into columns 40–69. A maximum of 360 subjects may be coded in this section using a single or direct punch per subject. A direct punch subject classification code in inorganic chemistry has been worked out by the author (Figure 14-12) with about seventy subjects. Five new subjects (on the average) are added every year. In columns 70–80 the small numbers above the row of zeros (or sevens) represent the first digit of the atomic number and the numbers in the vertical columns represent the second digit³⁷.

The file discussed next was developed at the U. S. Geological Survey, Washington, for the maintenance of a bibliography in geochemistry³⁸. Standard $6\frac{5}{8} \times 7\frac{1}{2}$ inch McBee punched cards with five holes per inch and a double row of perforations around the entire perimeter are used. The section in the upper right-hand corner is designated to code the author's name. The numerical breakdown permits the coding of only one author per card. The first and second letters of the name are broken down alphabetically into 99 subdivisions. Such classifications are available from the card manufacturers. The same number of chemical elements may be coded in a similar manner by punching the number corresponding to the atomic number of the element. Two fields of four holes numbered 7, 4, 2, 1 each are used in the upper margin of the card to code the year of publication. The right field is used for units and the left for tens. By limiting the use of each field to nine entries, it is possible to attain 99 entries in the two fields using a total of only eight holes. The hole marked "zero" is punched when 10, 20, 30, etc., are desired and also when 01, 02, 03, etc., are to be indicated. Simple numbers are distinguished from combined numbers by deep punching instead of shallow. The century is neglected inasmuch as most entries for this particular file are in the 20th century. The section marked "main element" at the top center of the card was established for

³⁷ Kirschner, S., "A Simple, Rapid System of Coding and Abstracting Chemical Literature Using Machine-Sorted Punched Cards," presented at the Atlantic City meeting of the American Chemical Society, Division of Chemical Literature, Fall 1956.

³⁸ Breger, I., "Design of Simple Punched Card Systems, with Reference to Geochemical Problems," accepted for publication in *Economic Geology*.

Col. No.	Numbers			
	1	2	3	4
40	Acids & Bases	Monodentate Ligands	Raman Spectroscopy	
41	Analytical Procedures	Magnetic Moments	Reaction Mechanisms	
42	Apparatus & Equipment	Mass Spectrometry	Research Proposals	
43	Bidentate Ligands	Microscopy	Resolution Procedures	
44	Biography	Microwave Spectroscopy	Review Articles	
45	Biological Chemistry	Nomenclature	Rotatory Dispersion	
46	Bond Types & Properties	Non-aqueous Solvents	Safety	
47	Chromatography	Nuclear Chemistry	Science Reviews	
48	Cyclopentadienides	Nuclear Magnetic Resonance Spec.	Special Compounds	
49	Dielectric Constants	Organometallic Compounds	Stability Constants	
50	Dipole Moments	Organosilicon Chemistry	Stereoisomerism	
51	Coordination Compounds	Organozirconium Chemistry	Theory	
52	Education	Oxidation-Reduction	The Trans Effect	
53	Electrolysis (not polarog.)	Patents	Ultraviolet Spectroscopy	
54	Emission Spectroscopy	Periodic System	Unusual Oxidation States	
55	Geometric Isomerism	Photochemistry	Visible Spectroscopy	
56	Hexadentate Ligands	Photometric Titrations	X-ray Spectroscopy	
57	History of Chemistry	Polarography	Zone Melting	
58	Industrial Chemistry	Poly-acids		
59	Infrared Spectroscopy	Polydentate Ligands		
60	Ion Exchange	Polymerization		
61	Kinetics	Preparations, Laboratory		
62	Laboratory Techniques	Properties, Chemical		
63	Lecture Demonstrations	Properties, Physical		
64	Molecular & Atomic Structure	Quantum Chemistry		

Figure 14-12. Direct punch subject classification code in inorganic chemistry.

the use of spectroscopists and analytical chemists. By punching the number 79 (the atomic number of gold) in this field, it can be shown that the abstract is concerned with an analytical technique for gold. Other elements, designated by appropriate punches on the side of the card, might be used to indicate interfering elements. A number was assigned to each file in order to avoid confusion between a number of files in the same office using different codes but the same card. This number is punched in a single field at the upper left of the card. The sides of the card carry both elemental and numerical designations for each hole. The bottom of the card has been left open for entries specific to any file being developed. Space has been reserved for a 100-subject breakdown in the lower corner of the card with the remainder set up for numerical coding should such be desired. Frequently recurring subjects are punched in the outer holes. The deep punch is reserved for those which occur less frequently. Often a reference appears in which naturally-occurring organic substances are related to various chemical elements—a situation in which it is desirable to use not only the subjects but also the elements direct-coded on the sides of the card. Although an attempt to apply two codes to the same series of holes sometimes leads to difficulty, it is possible to do so with a minimum of ambiguity. Should a reference occur for which it is necessary to code both subject and element on the sides of the card, it must be indicated that such an abnormal situation exists. Punching the hole as indicated in the lower left-hand corner of the illustration (Figure 14-13) shows that one or more elements are coded along the sides. Although a system such as this leads to the isolation of a number of cards which must then be sorted by hand, it has the great advantage of doubling the number of entries that can be made on the sides of the card. The hole marked "reprint" is punched to show that a copy of the paper referred to on the card is already available in the author's file. Each reprint is cemented into a separate folder and numbered and the number is noted on the related punched card.

In 1951 the National Association of Corrosion Engineers (NACE) initiated an Abstract Punch Card Service in which subscribers are provided with almost 2,000 coded corrosion abstracts per year. These are printed on punched cards which are pre-punched for subject matter by the NACE. The cards are so marked that the subscriber may punch them to indicate the author, journal reference and original reference date. The NACE abstract subject filing index used for the cards is divided into eight main topics: general, testing, characteristic corrosion phenomena, corrosive environments, preventive measures, materials of construction, equipment and industries. A 5 x 8 inch McBee Keysort card with a double row of holes around the perimeter is used. The holes in the outer row along the top of the card are numbered 1 through 28 and notched by the NACE

FILE	AUTHOR	YEAR	SUBJECT
AUTHOR Francis Owen Rice and Melvin Freamo			
REFERENCE Journal of the American Chemical Society, vol. 75, 548-9 (1953)			
TITLE The Formation of the Imine Radical in the Electrical Discharge			
(717)			

Figure 14-13. Keysort card used for the maintenance of a bibliography in geochemistry at the U. S. Geological Survey.

for the abstract subject. The inner row of holes along the bottom of the card is for the subscriber's use to designate journal reference and date. Both the outer and inner row of holes along the left-hand side of the card may be used to identify the authors of the reference³⁹.

Petroleum Industry

The Petroleum History Project at Northwestern University was established in 1954 by a grant from the American Petroleum Institute to prepare a complete history of the industry in America and an analysis of its effect on American life. The study is to culminate in 1959 with the publication of a two-volume history. To organize and standardize its collection of information and make it available to all members of the project, it was decided to use an edge-punched card system. This system is in a dynamic state: cards are being used constantly and subject interests vary from time to time. McBee Keysort cards were used with double rows of perfora-

³⁹ Mathay, W. L., and Hoxeng, R. B., "A Classification and Filing System for Corrosion Literature," *Corrosion*, 12 (11) 588-592 (Nov. 1956).

Petroleum History Project.

Subjects and Dates of Coverage

Top of card ("O" indicates Outer Row)

1 Bibliography	15 West Coast
01 Federal	015 Foreign (and export)
2 Biography	16
02 Inter-, intra-state, local	016
3	17 Exploration and drilling
03 Government document	017 Production
4	18 Geology, geography, history
04 Patent	018 Other sciences and technology
5	19 Refineries
05 Statistics	019 Refining and distillation
6	20
06 Court case, law suit	020
7	21 Railroad
07 Periodicals and serials	021 Pipe lines
8	22 Transportation (other)
08	022 Tidelands
9 Kerosine	023 Standard Oil Co.
09 Crude oil	24 Minor companies
10	024 Other major companies
010 Gasoline	25
11 Medical and other uses	025 Construction
011 Lubricants	26 Finances, earnings
1	026 Costs
012 Other petroleum products and fuels	27 Securities, speculation
13 Penna., Ohio, Allegheny, New York, West Virginia	027 Prices
013 Atlantic Coast (including Philadelphia, Penna.)	28 Estimates, resources
14 Mid-Continent, Gulf	028 Marketing
014 Rocky Mountains	29
	029 Supply

Left margin of card

30
31 Labor, employment
32 Safety, accidents, health
33
34 Insurance
35
36 Public relations
37
38 Waste, waste disposal, pollution
39 Conservation
40
41 Production rate, productivity
42
43 Illumination
44
45 Other related or competitive products and industries
46

Right Margin of card

47
48 Investigation (trust, government)
49 Regulation
50 Laws and legislation
51 Competition
52 Integration
53
54 Management
55 Associations and trade agreements
56
57 Social impact
58 Taxation
59
60 Research
61
62 Inspection, testing, standards, quality
63 Equipment and packaging
64 Storage

Bottom of card (inner row)

1 1921+	11 1886-1890	21 1936-1940
2 ...-1845	12 1891-1895	22 1941-1945
3 1846-1850	13 1896-1900	23 1946-1950
4 1851-1855	14 1901-1905	24 1951-1955
5 1856-1860	15 1906-1910	25 1956+
6 1861-1865	16 1911-1915	26 ...-1872
7 1866-1870	17 1916-1920	27 1873-1893
8 1871-1875	18 1921-1925	28 1894-1911
9 1876-1880	19 1926-1930	29 1912-1921
10 1881-1885	20 1931-1935	

Figure 14-14. Subject list used at Petroleum History Project at Northwestern University.

tions at the top and bottom and a single row at each side. Subjects are punched along the top and both sides of the card. The complete subject list is shown in Figure 14-14. Publication dates are punched into the card in two fields at the bottom, in the outer holes. In this system the publication date is not as important as the date of coverage. For instance, a paper published in 1920 may contain interesting figures on the industry during the years 1900–1915. The date of coverage is indicated by punching the appropriate hole in the inner row at the bottom of the card. The author's name is coded in three fields at the bottom of the card using the outer holes⁴⁰.

Since 1943 the American Petroleum Institute Research Project 44, at Carnegie Institute of Technology in Pittsburgh, has been concerned with collecting, analyzing, calculating, and compiling selected values of physical and thermodynamic properties and mass spectral data on hydrocarbons and related compounds. As of June 30, 1955, the tables of the API Research Project 44 cover 1400 different compounds and include more than 150,000 individual numerical entries. These are available on 45,430 IBM punched cards. The cards carry a two-line interpretation which permits the information on each card to be read at a glance. Each compound was assigned a number which is given together with the full name on a name card. The first three columns of the name card show the group number to which the compound belongs. Each class of compounds is assigned a number which is punched into columns 9, 10 and 11. Columns 4 to 8 and 12 through 24 are left blank on the name cards. Column 25 contains the card number. In most cases, only one name is given and the card number is "one." When compounds have two or more names, the names are punched on separate cards and the cards are numbered 1, 2, etc., in column 25. The name of the compound is punched into the card starting with column 27. The holes on the data cards are assigned as follows: 1–5, table number; 6, footnote; 7–8, year of latest reference of data; 9, 10 and 11, compound number; 12, state (gas, liquid or solid); 13–23, stoichiometric formula; 24–26, card number. The actual property values for each compound are punched in the cards starting with column 27⁴¹.

The Information Services Division at Ethyl Corp. in Detroit maintains a file of information on additives used in hydrocarbon or oxygenated-hydrocarbon fuels or in natural or synthetic lubricants. This information is recorded on Remington Rand punched cards and searches are made using the Remington Rand mechanical sorter. The file covers the

⁴⁰ Krull, A. R., "Punch Card System for the Petroleum Industry," *Petroleum Eng.*, E27–29, E32, E34 (March 1956).

⁴¹ Sherman, J. "Physical Data on Hydrocarbons," *Petroleum Refiner*, **32**, (10), 145–149 (Oct. 1953).

United States patents and Ethyl Corporation technical reports on the subject. The information found in patents and company reports is abstracted and all compounds and classes of compounds cited are listed. In addition, the functions of the additives and the petroleum products in which the additives are used are noted. The abstracts are digests or notations of those parts of the reference pertinent to the file. From these abstracts two punched-card files have been prepared, a subject file and an author (or patent assignee) file. One subject card is prepared for each compound mentioned in a reference. The first five columns may be punched with an abbreviation of the name of the country, in the case of patents, or with the year date in the case of all other types of literature. The next seven columns contain either the patent number or the accession number assigned to the abstract. For column 13, a special code is employed for various types of Ethyl Corporation material, to indicate whether the particular abstract deals with a formal report, correspondence, or laboratory test data.

Column 14 can be used as a guide to some special types of information, to indicate for example that the compound coded on the card is not an additive itself but that it is reacted with something else to produce an additive. The number of functional groups in the compound is punched in column 15. A code is punched in column 16 to indicate the elements contained in the compound. The area extending from columns 22 to 45 is devoted to the code which describes the compound cited as an additive. Columns 46 to 70 are used for the codes which then define the functions of the compound, such as antioxidant or antiknock agent. In column 80 is coded a notation of the type of reference in which the compound was found—journal articles, government material, Ethyl material, etc. Finally, columns 81 to 90 are punched to show the type of petroleum products to which the compound is added. Other areas of the card are not in use at present. A modification of the chemical code developed by the Chemical-Biological Coordination Center is used.

One author card is prepared for each author or inventor. The company to which the patent is assigned or which employs the author is also punched into the card. Auxiliary files include a master file of all materials consulted in the course of gathering information for the punched-card file, including the notation of non-pertinent references that have been checked. The punched cards are filed according to a rough classification of the compound types, based on the atomic components of the compounds. Thus, compounds containing carbon, hydrogen and oxygen are coded with "H" in column 16 and are filed together. Cards of different colors distinguish the number of structural groups in the compound. The author file is

duplicated and filed alphabetically by author and alphabetically by company⁴².

The Humble Oil and Refining Company, Baytown, Texas, uses 5 x 8 inch Keysort cards with a double row of perforations for a catalyst file. This file was adapted for use in a technical man's personal file for information on a wide variety of subject matter. There are 56 numbered holes available for indexing at the top and bottom of the card as well as letters and symbols on either side for more specific coding. A coding system using two code numbers per subject was selected. The numbers are assigned to the subject by means of a random number table. This allows therefore a total of 1,540 separate items to be indexed in the file and as many as ten items to be indexed satisfactorily on a given card. In 1954 this file was indexed for 150 major subject headings and contained about 600 cards which represented about 3,000 separate items. Three fields on each card are used to code the authors name, the first two letters of which are punched directly into the alphabetical index. The top and bottom of the card are devoted to the subject index and the left side to a formula index⁴³.

Hobbies

McBee 5 x 8 inch edge-punched cards have been used to great advantage by T. T. Hill of the Edwal Laboratories in Ringwood, Illinois, to prepare topical exhibits of stamp collections and to index any technical data about them of interest to the specialist. From one to ten stamps can be mounted on the face of a card with the name of the country and the date typed on the card. Technical and descriptive information is coded and the file can be hand-sorted for information on such things as perforation method, type of ink and paper, subject, historical connection and type of stamp⁴⁴.

A noteworthy example of the variety of uses made of hand-sorted punched card systems is an application to contract bridge. The Bridge Hand of the Month Club, Inc., supplies two decks of playing cards perforated and notched along the narrow edges with three holes in each position and ten positions along each end. Needle sorting each hole in one position separates the four predetermined hands. After a set of four hands is sorted, the hands are played and scored in the usual way. The result is then compared with

⁴² Graham, M. H.; Hildenbrand, B. S.; and Weil, B. H., "Indexing of Fuel and Lubricant Additives By Machine-sorted Punched Cards," presented before the American Chemical Society, Division of Chemical Literature, Dallas, Texas. April 11, 1956.

⁴³ Hoffmann, E. J., "Use of Punched Cards For Filing Technical Data," Humble Oil & Refining Co., Texas Chapter Bulletin, 5, (4) 10-16, (May 1954).

⁴⁴ Hill, Thomas T., "Stamp Collecting and Punched Cards," Private communication to J. W. Perry, August 11, 1953.

the recommended bidding and play discussed in the accompanying instruction book⁴⁵.

It has been suggested that punched cards are also admirably suited to filing and retrieval of interesting and instructive hands. Cards held in a given situation, bidding and play, plus comments and analyses, can be clipped from magazines or newspapers and pasted on the cards. These in turn can be coded according to the players or to the various special features of the bidding and play.

A suggestion from the Edwal Laboratories involves the use of punched cards for indexing photographic slides. The holes used for coding are punched directly along the edges of the slide itself. Ten classifications are possible with five holes punched at the top and five at the bottom of the slide. By turning the box upside down and removing the bottom instead of the top, the lower row of holes may be needled in the conventional way. The file may be searched for subject, date, locality, etc. A master code card indicates the meaning represented by each notch. For example, a code card which reads, "Glacier Park, 1940 T₃" shows that all pictures taken in that locality at that time are notched in the top of the slide, third hole from the left⁴⁶.

Another system for indexing photographic slides involves numbering each slide as it comes in from the processor and giving the same number to a corresponding punched card. A code of numbers is assigned to an alphabetically arranged list of subjects and these numbers are notched on the card as they apply to the slide it represents. For example, # 1 on the subject list might be architecture and any slide whose subject is architecture will be notched in the first hole on the card. Up to 19,999 subjects can be recorded on the upper as well as the lower edge of the card. The ends can be used to code the year in which the slide was made. On the face of the card is typed the location of the subject, the shutter settings and the date the picture was taken. This system uses 5 x 8 inch McBee cards with a single row of perforations along the edges⁴⁷.

Miscellaneous

Included here are some of the many interesting uses of punched cards which, because of their subject matter or approach, do not fit into the preceding groupings.

⁴⁵ This system has been devised by the Bridge Hand of the Month Club, Inc., 28-36 214th Place, Bayside, New York.

⁴⁶ Patton, A. R., "Punch Card Filing System for Your Slides," *The Camera*, 73 (1) 63, 130, (Jan. 1950).

⁴⁷ Davis, L. R., "Locate Your Slides and Negatives With This Punch Card File System," *U. S. Camera*, 16, (9) 68-69, (Sept. 1953).

Arthur D. Little, Inc. of Cambridge, Mass., has found two new applications for punched cards. One is a coded and punched file for all data collected on explosives from 1950 to 1955. The literature group of the company has also established a McBee card system to classify company personnel by educational background and experience, thus enabling more efficient use of their personnel on various research problems⁴⁸.

The New York Society of Electron Microscopists has issued a bibliography on Keysort cards which will keep abreast of the literature in all fields of electron microscopy. An outstanding feature of this bibliography is its ease of use. It is already coded and punched by the bibliographer. Articles containing multiple subjects are easily found by each subject as well as by author. The first issue covering the years 1950-52 (approximately 700 cards) was available at the time of the publication of this notice⁴⁹. Early publication of material for 1953 and quarterly publications on current literature thereafter are planned.

The National Intern Matching Program acts as a central clearing agency for hospitals seeking interns and students seeking internships. Each student submits a list ranking hospitals by preference. This information (about a 30,000 item cross-index) plus quotas for each hospital, forms the input data for the IBM 704 which performs the actual matching. The result of the 704 operation is a matching of preferences of the hospitals and students so that each student gets the hospital of his choice. This is determined by the way the hospital ranks him and its quota. The 704 thus analyzes 30,000 applications to approximately 800 hospitals which have been named by 7,000 students. The actual 704 running time of 1956 matching was 1 hour and 45 minutes⁵⁰.

The Massachusetts Institute of Technology has compiled a bibliography of all important world literature on coffee for the Coffee Brewing Institute, Inc. of New York using specially-printed edge-punched cards (Figure 14-15). E-Z hand-sorted cards are used to index uncoded information on the date of publication, source of reference, subject, author, title, publisher and abstract of the reference. Although the cards do not have to be kept in any specific order for sorting, they are given a number so that they can be arranged in order when a list is prepared for distribution. The year of publication is set up as a simple code which makes it possible to select 399 separate years. A hole marked "completed" in the upper right-hand corner is punched only after a final check has been made

⁴⁸ "Current Research and Development in Scientific Documentation," compiled July 1957 by the Office of Scientific Information, National Science Foundation.

⁴⁹ "Bibliography on Electron Microscopy," *Science*, 118, (3066), 378, (1953).

⁵⁰ Personal correspondence from Joan R. McJoynt, National Intern Matching Program, Inc., Chicago, Illinois, to B. L. Haksteen, July 8, 1957.

YEAR OF PUBLICATION										COFFEE LITERATURE SURVEY										COFFEE BREWING INSTITUTE, INC.										FOOD TECHNOLOGY DEPT. - M.I.T.										COPYRIGHT 1955 BY ERNEST E. LOCKHART																			
TO 17 18 19 20 21 22 23 24 25 26 27 28 29 30										COFFEE GENERAL										COFFEE BREWING										COFFEE ROASTING										COFFEE GRINDING										COFFEE PACKAGING									
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30										1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30										1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30										1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30																													
Black, J. W.										Freshly ground coffee and "blown" tins. The Analyst 51: 403-404. 1926.																																																	
A quantity of Costa Rica coffee was ground after keeping for 8 days from the roasting time, and the evolution of gas immediately determined. For 200 g. of coffee, 52 cc. of gas was collected in 1 hr., 90 cc. in 5 hrs., and 132 cc. in 48 hrs., and this result is regarded as typical.																																																											
--Br. Ch. Abr.																																																											
The evolution of gas from ground coffee is probably not due to the action of air on the coffee but is occasioned by the gradual elimination of gas from the coffee, which was evolved during the roasting process but held under pressure in the roasted bean. The amt. of gas evolved varies with the degree of grinding, the severity of roasting and the lapse of time.																																																											
--Ch. Abr.																																																											
Abstract in Zeitschrift fur Lebensmittel-Untersuchung und Forschung 61-62: 541. 1931.																																																											
PAGE--																																																											

Figure 14-15. Card used to index information on the world literature on coffee.

and the information punched is confirmed. Some 10,000 references were accumulated in two years covering the period 1925 to 1949, and for each individual year from 1950 through 1955. Work continues on the 1956 references and for the years prior to 1925⁵¹.

Daily copies of all crime reports in the Los Angeles Police Department are posted to various daily charts which are used to prepare periodic statistical reports. The crime reports are then coded according to division of records number, date of occurrence, date reported, location, type of crime, who, what and where attacked, means and object of attack, trademarks, and description. The coded items are then punched into two types of IBM cards; property loss cards and miscellaneous complaint cards (Figure 14-16). A card is prepared for each separate offense. The cards are used for the preparation of routine and special reports, and have been an invaluable aid in the analysis of *modus operandi* to identify suspects and to locate possible suspects already in custody for some other offense⁵².

Specially designed 8½ x 11 inch McBee Keysort cards (Figure 14-17) have been printed for preliminary studies of the occurrence and character of deep-sea diving accidents by the Navy. This card was selected because of the limited number of the eventual total sample—under 10,000. The

⁵¹ Lockhart, E. E., "A Card Punch Bibliography on Coffee," report submitted to J. W. Perry, May 14, 1956.

⁵² "Modus Operandi as Developed by the Los Angeles Police Department," a report prepared by the Statistics Unit Planning and Research Division, LAPD, December 1955.

CR No.	Officer	Report	Date of Report	Time	Location	Property	Value	Offense	Arrest	Disposition	Remarks	Signature	Date
000000	000000	000000	000000	000000	000000	000000	000000	000000	000000	000000	000000	000000	000000
111111	111111	111111	111111	111111	111111	111111	111111	111111	111111	111111	111111	111111	111111
222222	222222	222222	222222	222222	222222	222222	222222	222222	222222	222222	222222	222222	222222
333333	333333	333333	333333	333333	333333	333333	333333	333333	333333	333333	333333	333333	333333
444444	444444	444444	444444	444444	444444	444444	444444	444444	444444	444444	444444	444444	444444
555555	555555	555555	555555	555555	555555	555555	555555	555555	555555	555555	555555	555555	555555
666666	666666	666666	666666	666666	666666	666666	666666	666666	666666	666666	666666	666666	666666
777777	777777	777777	777777	777777	777777	777777	777777	777777	777777	777777	777777	777777	777777
888888	888888	888888	888888	888888	888888	888888	888888	888888	888888	888888	888888	888888	888888
999999	999999	999999	999999	999999	999999	999999	999999	999999	999999	999999	999999	999999	999999

L.A.P.D.—Property Loss M.O.—Feb. 1952

Figure 14-16. Punched cards used to record crime reports in the Los Angeles Police Department.

card has space for writing in the identity of the patient and for special notes⁵³.

Group feeding operations are increasing as is indicated by research surveys conducted in the field of agricultural economics. Nutritional research is also advancing, and the application of this knowledge and the development of a technique by which food nutrients may be rapidly and accurately calculated, may lead to the adaptation of manually operated marginal punched cards⁵⁴.

Dr. Mary K. Bloetjes, Professor of the Department of Institutional Management, New York State College of Home Economics, uses punched cards as a teaching aid in her course on Cost and Production Control.

⁵³ Personal communication from H. W. Gillen, Physiology Branch, U. S. Naval Medical and Research Laboratory, New London, Conn. to J. W. Perry, March 27, 1956.

⁵⁴ Bloetjes, Mary K., "Management Research in Food Service Operations," presented at the 37th Annual Meeting of the American Dietetic Association in Philadelphia, October 29, 1954.

The card is a standard 80-column punched card. It contains the following sections:

- PERSONAL DATA:** Includes Name (J. J. Gillan), Age (30), Sex (M), and other personal details.
- DIVE DATA:** Includes Date of Dive (10/10/55), Time of Dive (1400), Depth (100), and other dive-related information.
- SYMPTOMS:** Includes First Symptom (Nausea), Location of First Symptom (Head), and other symptoms.
- OTHER SYMPTOMS DURING DIVE:** Includes a list of symptoms such as Dizziness, Nausea, Vomiting, etc., with corresponding punch holes for recording.

Figure 14-17. Punched card designed for Navy to record deep-sea diving accidents.

This course is taught by the case problem method, using a series of eight menu items analyzed for the various factors indicated on the card such as form of purchase, type of food, condition of food, amount of or absence of waste, etc. The cards are direct-coded in order to facilitate teaching⁵⁵.

The Bureau of Aeronautics of the Navy has a service test under way for the ultimate elimination of blueprints of engineering drawings and the substitution of microfilm mounted in electric accounting machine (EAM) cards. The EAM cards are punched and interpreted for each exposure of microfilm, and will contain the drawing number, microfilm frame number, Federal supply code and model designation of equipment. This information will be repunched into Filmsort aperture cards which will be handled by standard punched-card procedures⁵⁶.

FACSI Incorporated, Deerfield, Illinois, has developed a unique reference system. FACSI is a code word that represents the group of words Fast Access (of) Coded Small Images. The system combines edge-punched cards, a code specially designed for nondestructive testing literature, and

⁵⁵ Personal communication to R. S. Casey from Dr. Mary Bloetjes, November 4, 1955 and November 16, 1955.

⁵⁶ "Military Specification. Microfilming Engineering Drawings and Related Data, Requirements For," MIL-M-18872 (Aer), June 13, 1955.

articles printed directly on the cards. Each film size (8 x 10½ inch) card contains a perfectly readable (reduced 6 to 1) image of a complete NDT article. All cards are pre-punched with the proper code and may be kept in random order. Every article published in the *Journal of the Society for Nondestructive Testing* has now been reproduced and stored on a specific card. The notched hole coding permits the extraction of any card in a few moments⁵⁷.

Since this chapter does not attempt to be exhaustive but merely indicative of the varied uses of punched cards, there were many interesting articles which came to our attention that have not been discussed in the preceding sections. These are listed in footnotes 58–69.

⁵⁷ Staats, H. N., "Data Extraction in Nondestructive Testing," *Nondestructive Testing* (Jan.–Feb. 1957).

⁵⁸ Lenihan, J. M. A., "Isotope Catalogue on Punched Cards (Edge-notched)," *Brit. J. Appl. Phys.* 3 (29) (1952) (*Nuclear Data*.)

⁵⁹ Way, K. "Data-type Abstracts," *Physics Today*, 10, 17–18, (1957) (*Nuclear Data*.)

⁶⁰ Schwabe, C. W. and Davis, L. R., "Marginal Punched Cards in Veterinary Research," *Am. J. Vet. Research* 15, (57) 634–638, (Oct. 1954) (*Biological Data*).

⁶¹ Gey, K. F.; Kalbe, H.; Schön, H.; and Stegemann, H. "Documentation of Physiological-Chemical Literature on Punched Cards." *Hoppe-Seyler's Z. physiol. Chem.*, 301, 70–77 (1955) (*Biological Data*.)

⁶² Reumuth, H., "The Indexing of Chemical Compounds. A Contribution to the Problem of Organization of the Literature," *Z. Angew. Chem.*, 41, 1204–7 (1928). (*Laboratory Records*.)

⁶³ Preliminary Report on Research in Progress in Scientific Documentation, compiled August 1956 by the Office of Scientific Information, National Science Foundation. Section on Monsanto Chemical Company. (*Laboratory Records*.)

⁶⁴ Kountz, R. R., "IBM Punch Card Data-control in Pilot Plant Operation," presented at the American Chemical Society Division of Water, Sewage and Sanitation, Fall 1952. (*Laboratory Records*.)

⁶⁵ Jones, W. S. and Butterfield, P. H., "A Technical Information Service Using Punched Cards for Indexing and Retrieval," presented at the American Chemical Society meeting in Minnesota, September 12, 1955. (*Technical Services*.)

⁶⁶ "Guide to NACE Corrosion Abstract Punch Card System With Appendix A, Sections I–VI," published by the National Association of Corrosion Engineers, Publication No. 51-6, June, 1951. (*Chemical Literature*.)

⁶⁷ Demer, L. J., "Bibliography of the Material Damping Field." WADC Technical Report 56-180, June 1956, Wright Air Development Center. (*Chemical Literature*.)

⁶⁸ Peakes, G. L., "The Unit Card System in the Indexing of Internal Technical Reports," Chapter 11, pp. 149–164 in "Progress Report in Chemical Literature Retrieval," edited by G. L. Peakes, A. Kent and J. W. Perry, New York, Interscience Publishers, Inc., 1957.

⁶⁹ Peakes, G. L., "Experience with the Unit Card System for Report Indexing," Chapter 19, pp. 306–327 in "Information Systems in Documentation," edited by J. H. Shera, A. Kent, and J. W. Perry, New York, Interscience Publishers, Inc., 1957.

Chapter 15

A CASE HISTORY OF A ZATOCODING INFORMATION RETRIEVAL SYSTEM

CLAUDE W. BRENNER

Allied Research Associates, Inc., Boston, Mass.

AND

CALVIN N. MOOERS

Zator Company, Cambridge, Mass.

The Problem

A rapidly growing collection of research reports presented an acute reference problem to Allied Research Associates, Inc., Boston, Massachusetts, in 1954. This organization of engineers and scientists, doing research, engineering, and development in the aeronautical and physical sciences, had been expanding since its beginning in 1951, with three engineers on its staff. At first, personal files of reports were sufficient for the company's information filing and retrieval needs. Later, project files were set up, and reports touching on the different projects were segregated into these files. As more contracts were undertaken, the engineering staff increased, and the rate of influx of technical reports and papers steadily mounted. By late 1954 the company had a staff of fifty, and the bulging files held about 3000 reports, with more coming in every day.

It was evident that the files would very soon become unmanageable. The first step was to turn to conventional library techniques, and a library school graduate was hired. It was to be her job to organize the company report collection so that it could be used easily by the engineers. It was hoped that she would be able to merge the catalog cards that had been received from ASTIA, AEC, and NACA. This required considerable knowledge of the subject matter and it was soon evident that in order to do the job extensive assistance would be required from the engineers. Not all of the incoming reports had catalog cards, and the analysis of their contents could not be left completely to the librarian. Moreover, it was the impression at the company that even had it been possible to interfile the various cards, the burden of filing multiple cards for each report would soon have become intolerable. The company therefore decided that a card catalog system would not be adequate to serve its needs.

Preliminaries to Setting up a Zatocoding System

At this point the engineer who supervised the library operation learned of the Zatocoding system. He and several other engineers witnessed a demonstration and tried sorting a pack of Zatocards. They also did some additional investigating. They contacted clients of the Zator Company to see how their systems had worked out. They studied the available books on punched cards. They got prices of other cards and equipment. They checked to see what assistance salesmen of other equipment could give in setting up a retrieval system for Allied Research's highly technical field. They estimated the likelihood that a conventional library system would be satisfactory. Probably the most decisive reason behind their final choice was the technical guidance provided by the Zator Company during installation of the system (Chapter 3).

The Zatocoding System

The Zatocoding system has three parts. There is the strictly mechanical part represented by the Zator "800" Selector and by the edge-notched Zatocards. This is the most tangible part of the system, though in some ways it is the least important. The second part is the technique of using random-like descriptor code patterns and of notching these code patterns into the edge of the card in superimposition. This is the Zatocoding technique. The third part is by far the most important and requires the most explanation. It is the system of "descriptors" by which documents are characterized, and by means of which retrieval questions are turned into prescriptions for search.

One card is made up for each of the reports in the collection. Notches along the edges of the cards permit a mechanical sorter to scan the cards and to select some of them. The subject content of each report is related to the pattern of notches in the card by the coding scheme. Therefore the sorter is able, by a strictly mechanical process, to select cards from a pack according to subject matter. All the cards are scanned for each retrieval question. Complete scanning has the advantage that the cards need not be kept in any order. Card filing is thus eliminated.

The Zator "800" Selector

Figure 15-1 shows the Zator "800" Selector¹ in operation. A pack of about 200 cards is placed in the black, box-like upper part of the selector. The box is vibrated by a small motor, as shown in Figure 15-2. Near the bottom of the box are rods or needles, which run from front to back. Each of the rings shown in the Figure is attached to a rod. It is easy to pull the

¹ Mooers, Calvin N., and Charlotte Davis Mooers, "Card Selecting Device," U.S. Patent No. 2,665,694 (1954).



Figure 15-1. Sorting cards with the Zator "800" Selector. Cards are taken from one of the side trays, are sorted, and then are placed in the other side tray. The accepted cards are dropped to the table in front of the machine.

rods out by means of the rings and to insert them again in a different selective pattern.

The Zatocards, like the one shown in Figure 15-4, have notches along the edges representing different subjects. In making a selection the pack is placed in the selector machine with one of the notched edges resting on the sorting rods. Most of the cards in the pack rest on the top of the grid formed by the rods. However, some of the cards, as shown by Figure 15-3, have notches in the position of each of the selector rods and are not supported on top of the grid. These shake down a little way from the rest of the pack, and are the desired cards.

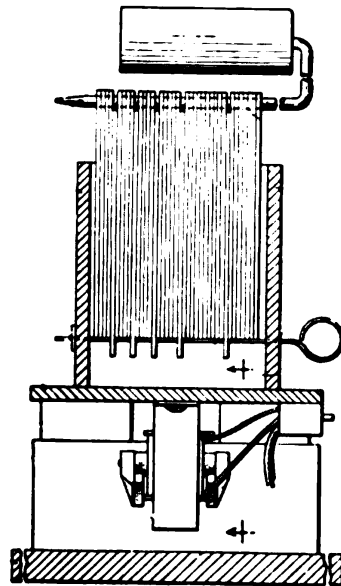


Figure 15-2. Cross-section of the Zator "800" Selector showing the vibrating motor and the manner in which most of the cards stay on top of the selecting rods.

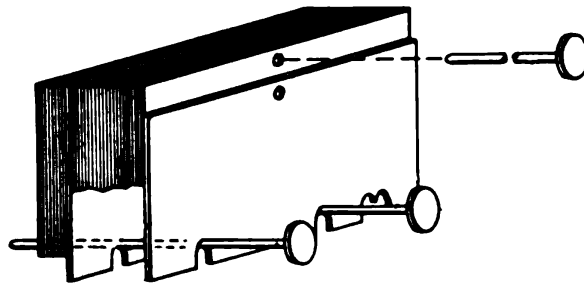


Figure 15-3. Diagram showing how the cards whose notches fit the pattern of the selector rods drop from the rest of the pack.

Looking again at Figure 15-3, it is seen that the pack of rejected cards can be engaged by a rod or tool inserted through the holes near the top edge of the cards. The desired cards, having dropped down a little, are not so engaged. Thus, when the tool is raised, the pack of rejected cards is held on it and lifted out of the selector. The desired cards are not engaged by the tool and drop free from the pack to the table top. For cards with coding on both the top and bottom edges, the selection operation for the codes on the second edge is carried out after the cards have been sorted according to the first edge.

About one second of the vibrating action of the selector is sufficient for complete separation. The speed attained in sorting depends on how nimble the operator is with his hands. Speeds of better than 800 cards per minute are easily attained.

Zatocards come in two styles. One style has notches only in a single edge. It has 40 notching sites. The card used at Allied Research has two edges given over to notches. This style has a total of 72 notching sites, and nearly twice as many descriptors can be notched into the double-edge cards. To sort the double-edge cards, the selector is first set up to scan the top edge of the cards. All the cards in the collection are run through the selector, which gives a partially selected pack of only a few hundred cards. The selector is then set up for the patterns on the bottom edge of the cards and the small pack of partially selected cards is run through. The second sorting goes very rapidly because there are at most only a few hundred cards involved. The cards that emerge from the second selection are the desired ones. Most of the selection time is taken up by the sorting on the first edge. For this reason, the speed of sorting is almost the same regardless of whether single-edge or double-edge cards are used.

Random Superimposed Codes

The second part of the Zatocoding system is the random superimposed coding method called Zatocoding^{2, 3, 4}. A pattern of notches is established for each subject covered by the documents in the file. The subjects to be coded are all overlapped or superimposed in an undivided area of the card. If two edges of the card are used, they are used as if they were one long edge. One might think that the superimposing would lead to an awful mix-up, but it doesn't, provided the code patterns do not resemble each other too closely. One way is to use random patterns; these are patterns generated by flipping a coin, or by some other similar means. The Zatocoding method teaches that any patterns that are "random-like" in the sense that the individual code marks are well scattered and fall with approximately equal incidence on all the coding sites can be used for coding. A list of random-like patterns has been prepared for Zatocoding systems to eliminate the necessity of deriving new patterns for each installation.

The Zatocoding method of using superimposed random-like code patterns is illustrated in Figure 15-4. The Zatocodes for the various descriptors have been written in on the card. The first two numbers represent notches

² Great Britain, Patented, No. 681,902, 3 September 1948; Canada, Patented, 1956, No. 534,926, 25 December 1956; U. S. Patent pending.

³ Mooers, C. N., "Zatocoding Applied to Mechanical Organization of Knowledge," *Am. Doc.*, **2**, 20-32 (1951).

⁴ Mooers, C. N., "Choice and Coding in Information Retrieval Systems," *Transactions of the Inst. of Radio Engineers Professional Group on Information Theory PGIT-4*, pp. 112-118 (September 1954).

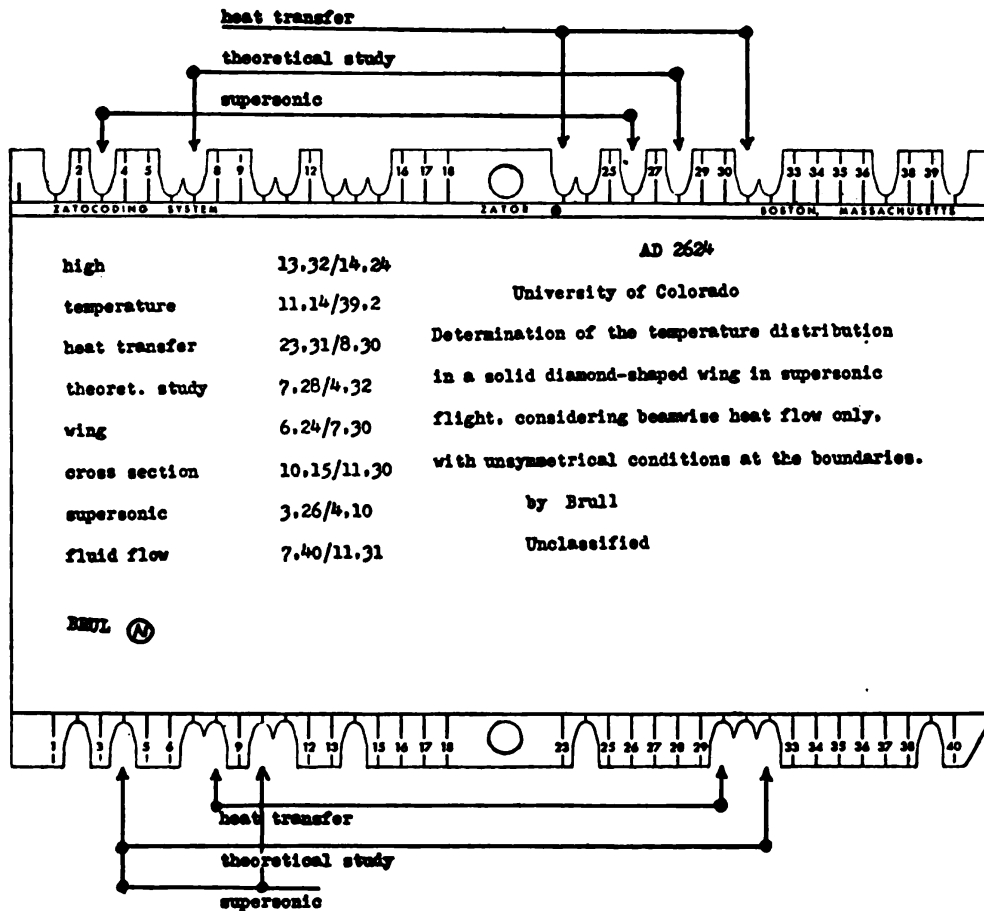


Figure 15-4. A Zato-card from the collection of the Allied Research Associates, Inc. showing descriptors and their Zato codes at left and the report file number at upper right. The lines and arrows illustrate how the three descriptors are coded.

in the top of the card; the second two, notches in the bottom. In the actual system, however, there is no need to write in the code numbers. Figure 15-4 also illustrates the manner in which selection is performed by Zato-coding. In the case shown, three subjects simultaneously define the desired selection. They are "heat transfer," "theoretical study," and "supersonic." The individual codes, and the way they are superimposed to form the total selective pattern, are shown by the diagram. The arrows correspond to the total selective pattern of rods set up in the card selector. Evidently selection will be made by this selective prescription, since there are notches in the card in every position where there is a selector rod. Note that a selected card may have more subjects (and thus more notches) than the selecting prescription. Selection extracts the cards that have at least all of the prescribed descriptors.

In addition to the handful of cards that contain the descriptors that

were prescribed, there will often be two or three cards on which the descriptors do not correspond in any way with the prescribed ones. These are the "extra cards" of Zatocoding. They are harmless, because they are so few in number and are so easy to discard. All the cards with the prescribed combination of descriptors are sorted out, so no cards are ever missing. The fraction of extra cards is mathematically predictable, and can be set to as small or large a value as may be desired by varying the number of notches per pattern in the Zatocodes.

The Descriptor Dictionary System

In contrast to the rather straight-forward mechanics of the code scheme and the selector is the third part of the Zatocoding system. This is the intellectual part and is called the descriptor dictionary system; it is the most important part of the system. It is called a descriptor dictionary system because it is not merely a list of subject words. Instead, it comprises several different kinds of lists, each having a definite function. It is the intellectual tool that couples the mind of the information searcher to the hardware of the Zatocoding system in such a manner that the hardware does the work of selecting the desired subject matter from the file.

TABLE 15-1. A PORTION OF THE ALPHABETICALLY ARRANGED SCOPE NOTES. DESCRIPTORS ARE PRECEDED BY ASTERISKS. TERMS NOT DESCRIPTORS [n.d.] ARE CROSS-REFERENCED TO DESCRIPTORS.

<i>*Stability</i>	135	12-11: 39-35
In aeronautical engineering, pertains to the study of aircraft stability as used in conjunction with *Static, *Dynamic, *Lateral, *Longitudinal. Also refers to instability, such as buckling or other structural instabilities. Use with *Derivatives in stability and control studies. For Lateral-longitudinal Stability Coupling (n.d.), use *Stability plus *Lateral plus *Longitudinal plus *Interference.		
<i>*Stall and Buffet</i>	136	38-16: 26-7
Stall pertains to the condition of partially or wholly separated flow on air-foils at high angles of attack. Buffet is the disturbance due to periodic boundary layer separation on a surface or the motion of a surface in a fluctuating wake.		
<i>*Static</i>	137	38-3: 28-25
With *Stability, pertains to static stability studies.		
<i>Statistical Mechanics</i> (n.d.)		
Use *Thermodynamics		
<i>Statistics</i> (n.d.)		
Use *Probability.		
<i>Stick Force</i> (n.d.)		
Use *Control plus *Biology.		
<i>Strain Gage</i> (n.d.)		
Use *Stress and Strain plus *Instrumentation.		
<i>*Stress and Strain</i>	138	15-3: 33-8
Any process involving the loading and deflection of structures, e.g., bending of beams, deflections of plates, theoretical elasticity studies, elastic behavior. Use also for Torsion (n.d.). With *Instrumentation it means Strain Gage (n.d.).		

A descriptor is something like a "subject heading" in library practice, though it is usually much broader in meaning. For instance, a subject heading might be "oils—effect of temperature on viscosity." In descriptor analysis, the separate descriptors "oil," "thermal," and "viscosity" would be used together to delineate this meaning. Each descriptor is a word-symbol standing for an idea or concept, generally of a rather broad scope. The particular scope of meaning for a descriptor is *assigned* in such a way that the descriptor will be most useful for retrieving information in a specified collection. Thus, the assignment of meanings at Allied Research is in part quite different from those assigned in other Zatocoding systems. Retrieval meanings need not conform strictly to standard technological usage of the word chosen to be the descriptor symbol. Because the meanings are often slightly different from the ordinary usage, it is essential that the descriptor dictionary system include a list of "scope notes," with a scope note for each descriptor. An alphabetically arranged list of scope notes such as shown in Table 15-1 then makes the full range of assigned meanings easily accessible to anyone desiring to use the Zatocoding system. These special descriptor meanings are private, for use in retrieval only, and there is no intent (nor likelihood) of imposing them upon ordinary speech or technical writing within or outside the company.

Deriving the Schedule of Descriptors

The schedule of descriptors is the most important component of the dictionary system. At Allied Research, a panel of four top engineers and physicists worked together in deriving a schedule of descriptors. With this panel of top personnel, problems of scientific and company policy as they affected the future use of the retrieval system could be settled on the spot. Thus, in areas where the company expected to embark on a new line of endeavor, the group was anxious to make sure that appropriate descriptors were obtained.

Deriving the descriptors is a strictly empirical process. On four separate occasions the Allied Research panel met with the Zator representative. A stack of reports, giving a typical sample of their file, was brought out and placed on top of the conference table. The top report was taken, its title and abstract were read to the group, and it was passed around for a brief examination of its contents. Then the question was posed, "Why would anyone at Allied Research be interested in using this report?" The answer may have been that it was about *propellers*, that it was about propeller *aerodynamics*, and that it was a *wind tunnel* study. Each of these was taken as a presumptive descriptor and written down. The same empirical process was followed with the next report, and so on. On more than one occasion, sad experience has given convincing proof that descriptors "dreamed up" in an armchair without reference to actual reports are worthless. This

TABLE 15-2. A PORTION (ABOUT ONE-QUARTER) OF THE DESCRIPTOR SCHEDULE AT ALLIED RESEARCH ASSOCIATES, INC., SHOWING THE GROUPING OF THE DESCRIPTORS AND THE WAY IN WHICH LEADING QUESTIONS ARE ASKED.

<i>What material was studied?</i>	<i>Is the process dynamic (rather than static)?</i>	<i>Are there specific aerodynamic loads?</i>	<i>Is structural strength and elasticity involved?</i>
Metals	Vibrations	Lift	Stress and strain
Gases	Transient response	Drag	Plasticity
Plastics	Impact	Moment	Failure
Aluminum	Stability	Gust	Ultimate properties
Magnesium	Velocity	Pressure	Material properties
Titanium		Center of application	Aeroelasticity
Air		<i>e.g., aerod. center, center of pressure, etc.</i>	Flutter
<i>What is the type of fluid flow?</i>	<i>Is it a stability and control problem?</i>	<i>Or is there another aerodynamics problem?</i>	<i>Is a thermal process involved?</i>
Fluid flow	Stability	Boundary layer	Thermodynamics
Internal flow	Control	Aeroelasticity	Thermodynamic constants
Subsonic	Static	Flutter	Combustion
Transonic	Dynamic = Trans. resp.	Downwash	Heat transfer
Supersonic	Longitudinal	Stall and buffet	Cooling
Hypersonic	Lateral	Interference	Convection
Laminar	Derivatives	Hydraulics	Conduction
Turbulence	Damping	Trajectory	Thermal
Slip flow	Weight and balance	Droplets	Radiation
Compressibility	<i>e.g., center of gravity, moments of inertia, etc.</i>	Modifying Technique	Aerodynamic heating
Viscosity		Performance	
Vortices			
Shock waves			
Finite span			

empirical procedure of discovering descriptors is surprisingly rapid. By the time that some fifty reports (selected to give an approximate cross-section of the company's interest) had been processed, more than 80 per cent of all the descriptors in the final schedule had been found.

In this stage of developing their system, the panel had many lively discussions about the theory and practice of using descriptors in information retrieval. These discussions were encouraged by the Zator representative, and the various points raised provided an excellent opportunity to bring up the experiences of other Zator clients who had similar problems.

At the second meeting of the panel, about a week after the first meeting, the descriptors obtained so far were written down on a large sheet. This was the first draft of the descriptor schedule (see Table 15-2). Related descriptors were grouped together in the draft; duplicate descriptors were eliminated.

Additional reports were then analyzed in the same way. Now the panel began to use the draft schedule as a guide. A few more descriptors were added, and rough spots in the draft schedule were ironed out. During this stage, scope notes were being written on index cards (for later typing in list form). Decisions made by members of the panel about the usage of the descriptors were thus written down while the problem was fresh in their minds. At various times the Zator representative would ask questions or offer criticisms to make sure that the panel was aware of the consequences of their decisions. Except for the teaching and guidance of the Zator representative, the panel did all the work in deriving their schedule of descriptors.

The panel at Allied Research put in a total of less than 150 man-hours from the beginning of the process until the schedule was ready to hand over to their clerical staff for typing and code assignment. This time included the "homework" that was assigned to the various panel members between visits of the Zator representative.

During the entire operation of deriving the descriptors, it was stressed that the primary orientation of a retrieval system must be toward the requirements of the user. One of the most important consequences of user orientation is that the descriptors must be broad in meaning. When the descriptors are broad, the user's intellectual universe can be covered by a relatively small list of descriptors. At Allied Research, 250 descriptors are used. Because there are so few descriptors in the system, they are relatively easy to remember, which is a definite advantage. The very breadth of meaning of each descriptor makes it easy to decide its applicability to a given document. Descriptors with finely drawn distinctions between them are avoided. Precision is not lost by using broad descriptors because ideas can always be synthesized by means of several descriptors. With so few

descriptors, it is easy to set them down on one big sheet, called the descriptor schedule. In this way, the analyst is able to see all the descriptors at once.

At Allied Research, the booklet containing the scope notes (alphabetically arranged and printed), and the large sheet which is the descriptor schedule, are distributed to the engineers who are most active in using the system or who are on the team analyzing the incoming reports. To aid further in finding the correct descriptors, the scope notes have interpolated words and expressions in ordinary technical usage with cross-references to the proper descriptor.

Analysis of the Incoming Documents—The Filtering Technique

When the incoming reports arrive at Allied Research's document center, they are given a preliminary screening to determine which engineer analyst is to handle each report. About sixteen engineers and scientists are on the analytical team. Each person gets the reports most closely related to his specialty. This procedure has the added advantage that it also keeps the specialists cognizant of the latest work in their fields.

The procedures adopted for document analysis are also user oriented. No attempt is made in analysis to code the message of the document by writing a little abstract using descriptor words. The descriptors and their codes are used for retrieval only, and the message itself will always be available in the document. Neither is there an attempt to secure pin point precision with the descriptors. Excessively narrow descriptors will only frustrate the user when he attempts retrieval.

The user of a retrieval system has a difficult problem. He is confronted by nothing but a schedule of descriptors supplemented by the scope notes. He is not sure what the file contains. He frequently knows nothing about the finer details in the reports. Thus, with only the schedule and scope notes, he must be able to formulate a prescription that will retrieve information, the nature of which may in large part be unknown to him. His success will depend largely upon how well the analysts originally did their job.

In conformity to the philosophy of user orientation, the document analyst is asked to place himself in the user's position. He does so in this way. First he reads or skims over the document. Then he lays the document aside and concentrates upon the descriptor schedule. He works down the schedule, descriptor by descriptor exactly as if it were a check list. For each descriptor he asks, "Would anyone at Allied Research who is interested in the content of this document use this descriptor as a part of his retrieval prescription?" or, "Does the meaning of this descriptor touch in any way upon the message of the document?" If the answer is "yes" to any of these, the descriptor is chosen as one of those to characterize the document.

This is known as a "filtering" technique and according to this technique, the schedule of descriptors is filtered through the message of the document. Those that remain in the filter are the chosen descriptors. If there is any doubt about the applicability of any descriptor it is resolved by choosing the descriptor. A doubtful descriptor may be just the one that will be tried in a retrieval prescription by some eventual user.

This technique has proved to be invaluable in giving the retrieval system a consistent intellectual structure. Consistency is a real problem. There are as many as sixteen or more contributing analysts at Allied Research and this group continues to change over the years. Yet their efforts accumulate in the form of the Zatocard collection. These cards must be consistent to be usable, and rigorous application of the filtering technique has forced internal consistency of the system.

The filtering technique also has another advantage. It does not require the analyst to have a highly technical background. If there were no filtering method, heavy demands would be placed on his ability and imagination. He would have to foresee all the possible uses of the document in order to decide which descriptors would apply. This is very difficult. However, the schedule of descriptors almost eliminates this problem because it serves as a check list of present and future contingencies as worked out by the top people in the laboratory. When the analyst uses the schedule as a check list, he only has to make very simple decisions.

The burden of using a schedule of 250 descriptors is eased by a simple process. About one-quarter of Allied Research's descriptor schedule is shown in Table 15-2. The descriptors are grouped, with each group being composed of similar descriptors. At the top of each of the groups there is a question, such as, "Is there a type of fluid flow?" In using this kind of a schedule, the analyst first looks at the questions. If the answer to any of them is "yes," then he picks out the one or more appropriate descriptors below the question. If the answer is "no," he continues to the next question. Use of the filtering technique in the Zatocoding dictionary system involves going through a list of about 20 questions rather than through 250 individual descriptors. Carefully chosen "leading" questions, as in this example, make the incoming document analysis particularly easy.

This grouping of descriptors is not a scheme of hierarchal classification. There are no generic or specific terms. Any descriptor can be used with any other, and more than one descriptor from a single group can be used to characterize a document. A typical document in Allied Research's collection has from six to fifteen descriptors in its characterization. It is sometimes convenient to place the same descriptor in two different groups. This is useful for a few of the descriptors that may appear in widely differing contexts. An alphabetically arranged list of descriptors is specifically not

used for the analysis of documents because of its inferiority to the grouped arrangement in providing accurate analysis.

An actual analysis proceeds as follows. The first decision of the analyst is whether or not to include the document. Obviously worthless material must not be allowed to increase costs or to dilute the system. By the time the analyst sees the document, it has passed this threshold of utility. The analyst then skims or reads the document. Depending upon the obscurity of the writing, or the richness of the content (there is often an inverse correlation), this usually takes from 5 to 25 minutes. Fifteen minutes is not a pessimistic average for technical reports. The analyst then takes the descriptor schedule and reviews the check list of questions, writing down the chosen descriptor words on a Zatocard. This step of filtering and writing down the descriptors takes about two minutes. The card then goes to the clerical staff who types or writes the title, authors' names, report file number, and any similar information on it. To save time and expense, the abstracts are not typed in. The clerical staff then marks the cards and notches them with the descriptor codes. The finished cards are not kept in any particular order. The documents are filed by number, and the process is complete.

Most of the analyst's time is taken in becoming familiar with the document. The Zatocoding System is not unusual in this respect. Regardless of the system used, comparable time will be required if the documents are to be analyzed to a like degree. The assimilation step accounts for 50 to 75 per cent of the total cost of operating a retrieval system, with clerical costs and overhead accounting for the rest.

Since the card services of ASTIA, AEC, and NACA are available to Allied Research for a large fraction of their reports, they are able to use a "clip and paste technique" to save typing. The stock used for Zato cards is heavy enough to support the weight of pasted-on material and it does not interfere with the sorting. Everything except the citation and the abstract is trimmed off the catalog cards before pasting. When such a Zatocard is sorted out, it carries the full printed abstract of the report.

Coding the Cards Accurately

Since Zatocoding uses random-like code patterns for the descriptors, and since random patterns are difficult to remember and to transfer accurately, there is a serious problem in coding. Its complete solution lies in the elimination of the mental transfer step. The technique is illustrated in Figure 15-5 which shows one page of a code pattern dictionary with a card in position for coding.

To use this code pattern dictionary, the clerk reads a descriptor from the card, finds the page and line of the descriptor, and lays the card down on the

detonation	(42)	35. 13 / 7. 40
differential equation	(43)	31. 40 / 27. 1
diffusion	(129)	24. 17 / 7. 16
downwash	(45)	23. 2 / 11.12

ZATOCODING SYSTEM
ZATOC
POSITION, MASSACHUSETTS

DOWNWASH

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

^
^
^
^

Figure 15-5. The code pattern dictionary in use for transferring the code pattern for the descriptor "downwash" to the Zatocard. With the card in this one position, the location of the code notches for both the top and bottom edges are marked with a pencil. The cards are punched later.

page under the descriptor entry. The first notching position of the card (site number one) is aligned with the vertical index line on the page. With the card in this position, the V-shaped marks on the page indicate the exact locations at the top and bottom of the card that are to be notched. The clerk transfers the code positions to the card with pencil marks. There is no error-prone mental step, so the accuracy is high. After the codes for all the

descriptors have been made, the marked sites are notched with a hand "ticket punch."

Scatter Coding for Author and Company Names

The last component of the dictionary system deals with authors' names, company names, and the like. Instead of setting these up as individual descriptors and explicitly assigning them code patterns, a technique of scatter code ciphering is used to produce random-like patterns. By using such a ciphering process, a vast number of assigned patterns for little-used names is avoided. Scatter coding is primarily used for author names. When there is more than one author, the additional authors are ciphered into the card together with the first.

As shown in Figure 15-6, a card to be coded with an author's name is laid down on the scatter code sheet with the left-hand edge of the card opposite the index "N" (for name). The first four letters of the surname are then spelled out. Two letters from the name are ciphered at the top of the card and two at the bottom. Notice that the letters of the two alphabets at each edge are displaced so that letters of high frequency like "e" or "t" do not coincide at the same site on the card. This kind of displacement of the alphabet insures that the scatter-coded entries will have notches with an approximate uniformity of incidence of notches across the edge of the card as required by the Zatocoding method. The scatter codes are suffi-

SCATTER CODING SHEET

		I																										II																									
		ABCDEFGHIJKLMN O P Q R S T																										S T U V W X Y Z 1 2 3 4 5 6 7 8 9 0 A B																									
		UVWXYZ 1 2 3 4 5 6 7 8 9 0 A B C D																										C D E F G H I J K L M N O P Q R S T U V																									
Publishing or Originating Agency	(C)																																																				
	(N)																																																				
Author Name	--																																																				
Specific Aircraft or Missile Test Operations	(S)																																																				
		AUTHOR: LARKIN LARK (N)																																																			
		III																										IV																									
		ABCDEFGHIJKLMN O P Q R S T																										S T U V W X Y Z 1 2 3 4 5 6 7 8 9 0 A B																									
		UVWXYZ 1 2 3 4 5 6 7 8 9 0 A B C D																										C D E F G H I J K L M N O P Q R S T U V																									

Figure 15-6. Diagram illustrating the ciphering of an author's name by scatter coding.

ciently random-like so that the usual number of names on any card will not upset the sorting statistics. Standard rules are employed with company names to eliminate parts of the name useless for sorting, like "corporation." Such rules enable the elimination to be done in a consistent and repeatable fashion. This particular version of scatter coding was chosen for use at Allied Research after several other schemes had been considered.

The Need for Rapid Cyclic Search—Machine Feedback of Information

The Zatocoding system at Allied Research is actively used by the engineers for a variety of reference problems. Despite their excellent familiarity with the system, and their background in the various technical fields covered, it frequently happens that an engineer making a search is unable at first to prescribe exactly what he wants from the document file. He has to do some browsing before he knows how to sharpen his question so as to obtain the best answer. The way he does this is to formulate the best trial prescription that he can. He then sorts the cards and looks over the selections. He goes through the titles and abstracts to see how close he came to what he thought he wanted. On the basis of this preliminary work, several things may happen. He may find exactly what he wanted in the way of information, or he may actually change his mind as to what he does want. A third possibility is that he may decide how better to prescribe his search. Thus he may omit some of the descriptors from his prescription and add a few others. With such a new search prescription, he is ready to make another search. He may make a second or even a third search. From each search he learns more about the content of the file and how to ask his questions to match his technical problem.

A cyclic search process is unavoidable in creative science and engineering ✓ because the questions that arise are often diffuse and the details of the looked-for facts and theories are at first unknown to the searcher. This is why a recourse to more elaborate coding or information analysis cannot eliminate this problem. The shortcoming lies in the user and not the retrieval system, and it is one of the functions of the system to educate the user at the question-asking stage. To do so, there must be a rapid feed back of corrective information. At Allied Research the cards can be completely searched in six or seven minutes, and they will provide an immediate answer in the way of titles or abstract.

In many cases particularly with diffuse questions, the occurrence of extra cards is helpful to the user. For instance, in making a selection from descriptors A, B, and C a few extra cards will often come from the selector with only descriptors AB, or AC, or BC of the original prescription. One descriptor is missing. These few extra cards have a valuable property in

that they give a random cross-section of information outside the original search prescription, but a selection that is still heavily biased in the direction of the search prescription. These cards have been called "subject induced extra cards."

As the cards drop from the selector, they have the title and often the abstract of the report in plain language. The cards also have all their descriptors written in. Thus when an interesting technical lead appears, it is immediately evident which descriptors should be used in following it.

An Evaluation

The Zatocoding System at Allied Research Associates, Inc. has been in operation for nearly four years (August 1958), and contains about 8,500 documents. The whole file can be searched in twenty minutes. New documents are added at a rate of about 150 per month. Searches are made on the average of once a day. Any engineer, after a short training period, can analyze documents and make searches. Two young women are working full time running the document center and coding and notching the cards. The system at Allied Research is generally accepted as a straight-forward working tool, quite in the same way that a desk computing machine is accepted and used. In seeking contracts, the company now stresses its outstanding ability to retrieve information.

The system has worked smoothly and achieved the expected performance. During the period of installation, the advice of the Zator representative was helpful in foreseeing pitfalls in advance. There was no substantial backtracking nor need to correct mistakes.

As a result of experience, about a dozen descriptors were added to the dictionary during the first half-year, and about the same number were dropped. Since then the coding dictionary has had considerable stability.

The costs for this commercial system are: \$45.00 monthly rental and license for the Zator card sorter and for the Zatocoding technique, \$15.00 per thousand for cards, a professional fee for assistance in installation, and traveling expenses incurred by the Zator representative.

In just a year, the technical files were smoothly transformed into an efficient operation which is an engineering asset to the company.

CHAPTER 16

THE USE OF PUNCHED CARDS IN LINGUISTIC ANALYSIS

REV. ROBERTO BUSA, S.J.

*Centro per l'Automazione dell'Analisi Letteraria dell'Aloisianum
Gallarate, Italy*

A chapter describing the application of punched cards to studies in the humanities has a place in a book such as this, because such application in part parallels, and in part coincides fully with its corresponding use in scientific documentation and in libraries. I maintain that it is mutually advantageous to consider how the same tool for investigation responds to the demands of many problems differing in their nature.

I am concerned here with "linguistic analysis" in a broad sense, rather than in any of the specific meanings that different schools have sought to impose upon the phrase. I refer to any type of investigation of language, whatever significance the word "language" can assume. For example, I include the study of phonetics, of glottology, grammar, or style. In a word, I speak of philology in its broadest sense, and of psychology. I speak only of the investigation of written material, or more strictly, printed words. Even studies of phonetics can be based on printed texts. Hence, I am not concerned with those other analyses dealing directly with human sounds, such as those conducted at the Haskins Laboratories¹ in New York, nor those conducted with devices on which data are not recorded in letters or symbols (e.g., studies of comparative phonetics).

The analysis of language is as old as the knowledge of human knowledge.

¹ These studies are aimed primarily at isolating the significant signals embedded in the speech stream and in analyzing their perception by the human listener. An analysis-synthesis technique is used wherein the speech is converted into visible patterns, the patterns are re-drawn in simplified form and, finally, the modified patterns are re-converted into synthetic speech to provide the acoustic stimuli for perceptual studies.

"Some Results of Research on Speech Perception," A. M. Liberman, *The J. Acoust. Soc. Amer.*, 29, No. 1, 117-123 (1957).

"The Interconversion of Audible And Visible Patterns As A Basis For Research In The Perception Of Speech," F. S. Cooper, A. M. Liberman, and J. M. Borst, *Proc. Natl. Acad. Sci.*, 37, No. 5, 318-325 (1951).

"Some Experiments on the Perception of Synthetic Speech Sounds," F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst, and L. J. Gerstman, *The J. Acoust. Soc. Amer.*, 24, No. 6, 597-606 (1952).

Even without disturbing Plato in his Dialogues, it would be necessary only to recall to mind the rhyming dictionaries and the hundreds of concordances that have been published since the invention of printing. In more recent times, there has been increased interest in literary statistics. Reference was made to quantitative statistical analyses in formulating psychological and stylistic laws, for example, on the length of phrases, on the distribution of phonetic accents, on absolute and comparative frequency of words, of parts of speech, or of phonemes (in the meaning of letters of the alphabet). There are scholars in the United States who have made important progress in this field.²

Therefore the subject matter to be analyzed is made up entirely of what can be transcribed from human speech into characteristic signs or symbols. It can be considered as having three levels. First of all, the word is the fundamental unit, and it is at the same time the graphic and semantic unit. Then there are sentences and phrases composed of more than one word. On the other hand, there are elements of each word, such as roots, prefixes and suffixes. In the same way we speak of atoms, molecules, and electrons.

Thus it would be interesting to know which are the words used by a person or an epoch or a language. How many are there? To what radicals can they be reduced? What is their frequency? their length? What are the rhythms of their accents? How are words distributed in phrases? What are the fundamental structures common to the phrases? There are many such questions.

The problem requires the searching, separating, arranging, correlating and study of a large number of small elements, tens of thousands of words, hundreds of thousands of letters. Such investigations must be repeated many times on the same material with different emphases and for diverse purposes.

For these studies we must record every unit of information on a free and manageable medium, such as a card. Punched cards permit multiple coding of the same information, and they can be sorted and re-sorted rapidly. In addition, the great—even enormous—quantity of cards to be handled, and the possibility of making automatic printouts directly from the cards, dictated the choice of machine-sorted punched cards. Among these I have finally chosen the IBM system, not only because Providence obtained for me the full collaboration of this company, but also because of the great flexibility of the system, because of the developments the company foresees

² For a list of U. S. scientists in this field, see for example Guiraud, Pierre. *Bibliographie critique de la statistique linguistique*. Révisée et complétée par Thomas D. Houchin, Jaan Puhvel, et Calvert W. Watkins, sous la direction de Joshua Whatmough. Utrecht, Editions Spectrum, 1954.

xix, 121 p. (Comité international permanent de linguistes. Publications du Comité de la statistique linguistique, 2).

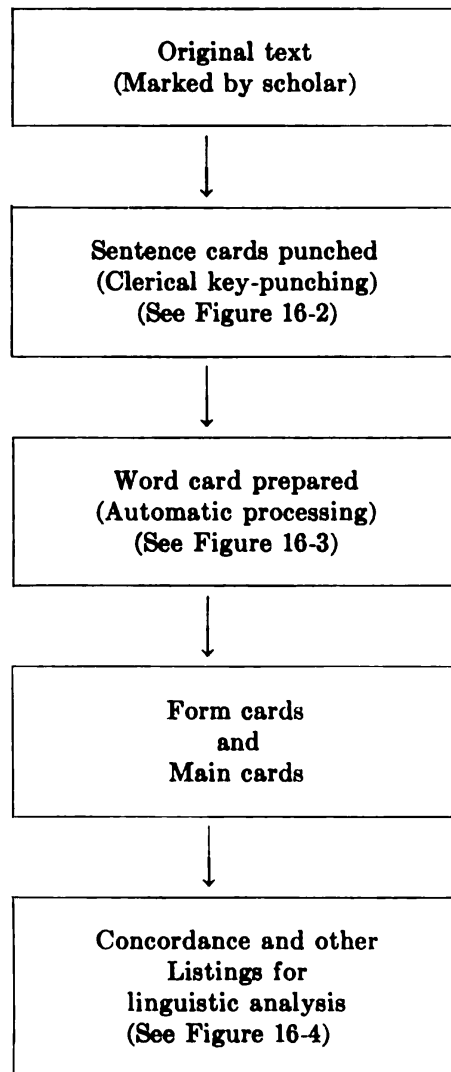


Figure 16-1. Summary of operations.

in the near future, and because IBM has been developing machine methods for scientific documentation.

I will now recount all that I have done, and all that there is still to do. I will use a flow chart (Figure 16-1) as the basis for my discussion. This was first prepared at IBM in Milan by Mr. C. Folpini and then completed in the offices of IBM in New York with the assistance of Mr. P. Tasman.³ It

³ "Literary Data Processing," P. Tasman, *IBM J. Research & Development*, 1, No. 3, 249-256 (1957);

"Literature and Document Research Automation," P. Tasman, *Automation Systems*, Engineering Publishers Division of The AC Book Company, Inc., 1958; 61-72.

will be evident that the process so described could be shortened considerably, if it is sufficient to obtain simpler results. However, it is useful to give the whole picture to show how much can be obtained if desirable.

Analysis of Words in a Text

I will concern myself here with the principal task, in terms of size and value, of linguistic investigation, which is making a concordance of a continuous text. It will not be difficult to apply these techniques to other problems, such as analyzing the answers of questionnaires, or the words found as items in a glossary.

The scholar marks the text to indicate how it should be recorded on the cards, noting the beginning and end of paragraphs and of sentences with their appropriate references. Also he distinguishes words quoted by the author from other writers, from the author's own words, etc.

Where it is important not to mark directly on the text so as to deface it, a sheet of cellophane may be placed over the page and appropriate markings made on it with washable inks.

Each line from the text is punched into a card, one line after the other, each with identifying reference to its place in the text. The maximum number of columns available for this punching is determined in advance, depending on the format wanted for the concordance. Thus a maximum sentence length is established. Words are never split between cards; rather, a word is started on a new card if it will not fit on the preceding one. See Figure 16-2.

The problem of verifying the punching is an important one, because an undetected error will always be repeated. Errors can be detected in the usual way by checking the card on the verifier, or by proofreading cards on which the punching has been interpreted.

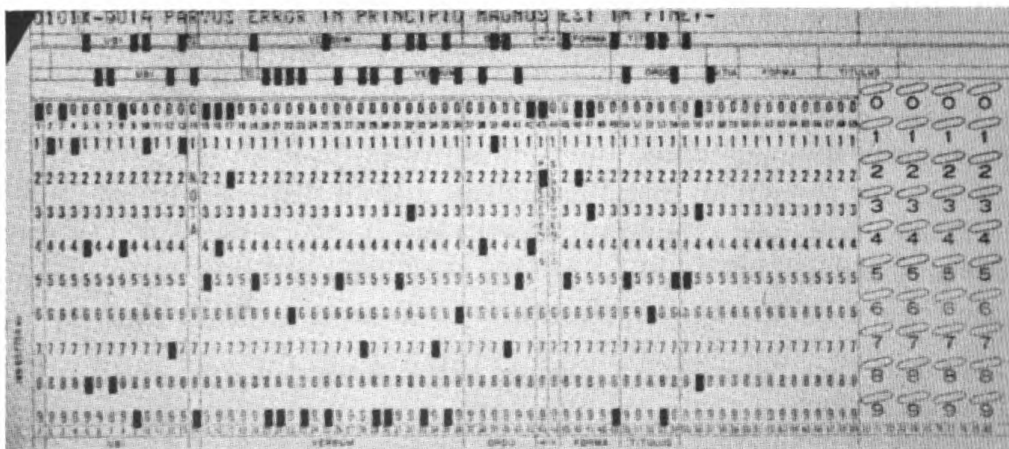


Figure 16-2. Sentence card.

These operations produce the first or fundamental group of cards called the *text cards*, or on the flow chart, the *sentence cards*. With this first and only data transcription, it is possible to accomplish mechanically, speedily, and accurately all of the most diverse and complex analyses.

To divide the sentences into single words, each on a separate card, the following methods could be used:

- (a) punching each word from the text onto a separate card.
- (b) simultaneous use of the sorter and reproducer equipment as described in the small volume, *Varia Specimina Concordantiarum*.⁴
- (c) using the Cardatype, recently developed by IBM.

The use of the Cardatype offers the advantage of preparing typed copy of the text while punching the individual word cards. Thus the context of the word is printed on the reverse side of each card.

This operation results in a second set of cards, the *word cards*. Each word is accompanied by reference to its place in the text. This file contains as many cards as there are words in the text. See Figure 16-3.

The word cards are alphabetized, using the sorter. Mechanical alphabetizing requires two passes of the cards through the machine for each column sorted. Thus sorting 100,000 cards containing words of 10 letters means, in effect, passing 2,000,000 cards through the machine. Depending on the model sorter used, from 30,000 to 60,000 cards per hour can be sorted. Therefore it could take from 35 to 65 hours, approximately, to accomplish the alphabetization. In other words, the machine would alphabetize from 1,500 to 3,000 words of 10 letters in an hour.

The operation can be shortened by several means, e.g., by separating first the shortest (one- and two-letter) words, then the next shortest words, and so forth. The shorter sets so divided are then inserted into the alphabetically sorted sets of longer words. The final result will be that all the words of the text are alphabetized and all identical words are grouped together.

Each group of identical words is given the same sequence number.

The accounting machine is used to print a list of all the words from the word cards. It is possible to prepare an abridged list on which only the different words appear. The machine will also print the total number of cards on which each different word appears. This gives the frequencies with which each different word appears in the text. The accounting machine may also be set to print the code number identifying the different forms of individual words.

When the summary punch is connected to the accounting machine, a third series of cards can be obtained while the list described above is being printed. These cards, called *form cards* or *different word cards*, contain each

⁴ Roberto Busa, *Varia Specimina Concordantiarum*, Fratelli Bocca, Milano, 1951, 180 pp.

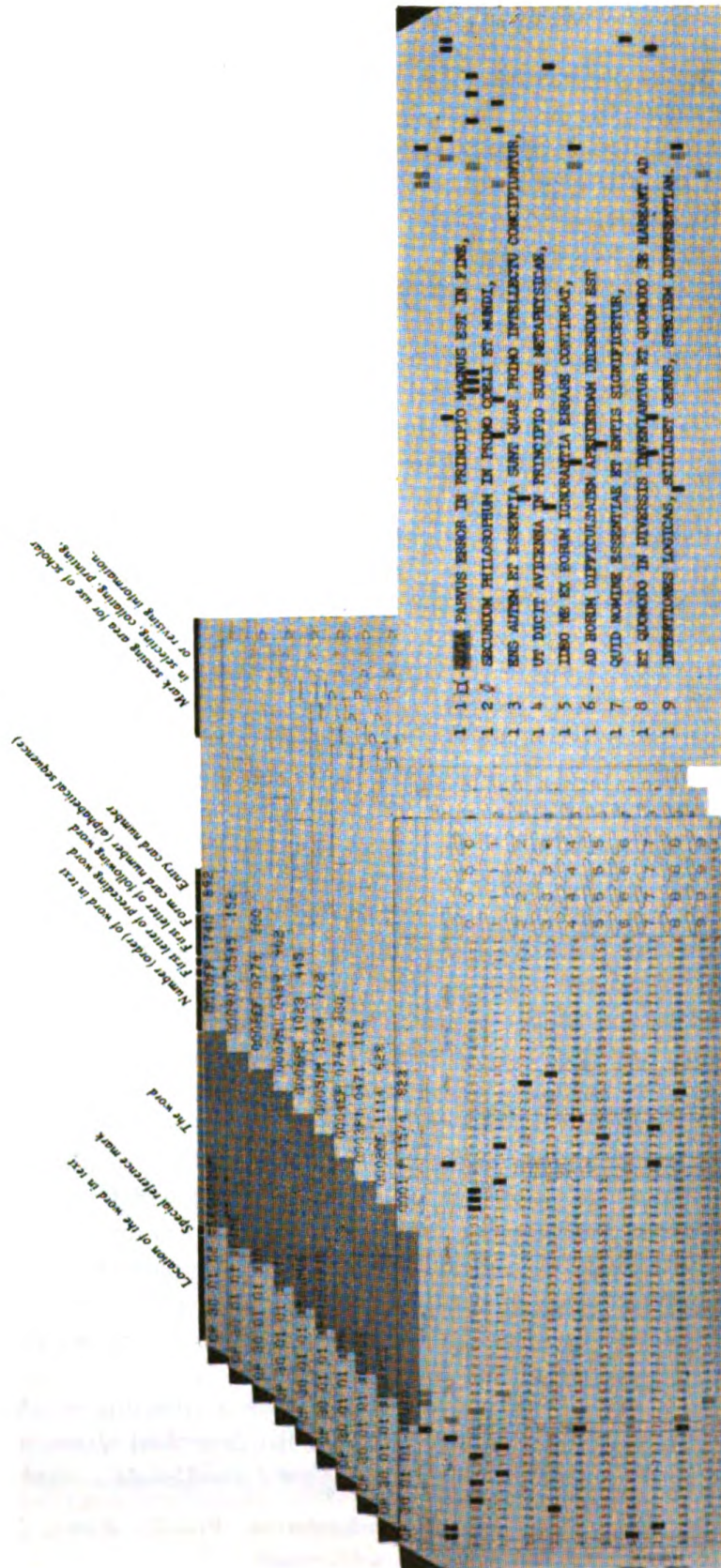


Figure 16-3. Example set of word cards.

different word with a number indicating its position in the alphabetic sequence and the total frequency of its appearance. Such a group of cards is not necessary for ordinary concordance work but it may open up the possibility of different and new investigations. Such cards, in fact, contain the summary of the author's vocabulary and can be analyzed indefinitely. Mark sensing techniques are particularly useful for this (as will be discussed later).

In the list of words described above, the words are considered according to their graphic structure. Therefore the scholars must separate cases of homographs, which turn out to be quite frequent; dismember words that comprise prefixes and suffixes each having a proper function also when isolated (such words may be considered as two words rather than one); assemble the separate words that are in reality just one verb form; and finally, regroup under the functional semantic unit all the diverse forms a word assumes according to case, tense, mode, etc.

Such work requires the competent responsibility of the scholar and it cannot be accomplished by machine. However, once such classification has been made, mechanical recognition of different forms of the same word could follow.

The *main words* must be punched one per card with a special layout designed to accomplish the functions of these cards. They must also be arranged in alphabetical order and numbered progressively. Such a number-code may be added to the cards of the other two sets, word cards and form cards.

In the form that I have summarized there is obtained from one initial punching of the text, four groups of cards. They are: the text cards and the word cards, that contain all the words of the text and represent two new editions of the entire text; and the form cards and the main cards, which constitute two summary indexes of the vocabulary used in the text. The first lists the words grouped according to graphic form, the other lists the same words arranged according to graphic-semantic units.

Note that the word cards are accompanied by the elements necessary to characterize their individuality. The dissociation of the text into its first elements is, therefore, entirely reversible: it is always possible to reconstitute the text from these elements. Such proper determinations, reserved and exclusive of each single word, are its various codes. In fact, every word is coded as to its location with the reference and with the number of its position in the text; it is coded as a morphologic unit with the progressive number that it acquires in the first alphabetic sequence; it is coded as a semantic unit, with the progressive number that it has in the last alphabetical order.

Besides, it is accompanied by its context. Such context may be punched or printed. It may be punched on the same card on which the word is

punched; a condition, however, that is restricted to about 50 to 60 letters and spaces, i.e., columns on the cards. It may be punched on another card, and then occupy even 70 or more columns, according to the length of the reference; such a card would then be the same text card. The context may be also printed in the spaces between the punched holes, and then it can be extended to twelve lines, and contain from 80 to 120 words; ample context that would almost always be sufficient to individualize the significance of the word without requiring the scholar to make frequent recourses to the printed text.

Finally, every single word can be accompanied by the first or last letter of the preceding word, by the first or last letter of the following word, besides the preceding quotation and the following quotation.

There also remains in the word card or in the form card or in the main card sufficient space for additional classifications to be applied, for example, manually by means of mark sensing. The researcher can make a symbol that tells what part of speech the word is, on what syllable the tonic accent goes, what is its length in letters or in syllables, and other things, too.

So the resolution of the text in its first elements is completed. The four groups of cards represent the material suitable for whatever investigation in whatever direction: investigation that, in its quantitative aspect involving large numbers of small elements, is accelerated enormously by mechanization and rendered more accurate, more certain, and absolutely complete. The same cards will serve the most diverse analyses, because once the cards are selected according to a determined order, they may be brought back rapidly to their first order and subjected to new research.

One can obtain, from the one and initial punching of the text, various listings as summarized below and exemplified in Figure 16-4.

(1) The general catalog of vocabulary of the author, richer in prerogatives and more abundant of context than the same monumental concordance of *TLL*⁵ prepared in Munich, Bavaria.

(2) The listing of the cards in various forms, using the accounting machine at a rate of 4500 to 9000 lines/hour. For example: (a) The text cards may give a reprinting of the entire text. (b) The word cards may give the

⁵ *TLL*—*Thesaurus linguae latinae*, editus auctoritate et consilio Academicarum quinque Germanicarum Berolinensis, Gottingensis, Lipsiensis, Monacensis, Vindobonensis. Lipsiae, Teubner, 1900.

Current volumes read: *Thesaurus linguae latinae*, editus iussu et auctoritate consilii ab academiis societatisque diversarum nationum electi.

"The great dictionary of the language, in Latin, indispensable in the university or large reference library. Plans to record, with representative quotations from each author, every word in the text of each Latin author down to the Antonines, with a selection of important passages from the works of all writers to the seventh century." Winchell, Constance M. *Guide to reference books*. 7th ed.

LATERCULUM VERBORUM

Numerorum qui singula subsequuntur verga primus frequentiam, alter cui lemmati in Rationario adunetur indicabit.

1 A	2	1
2 AB	1	1
3 ACCIPITE	1	2
4 AD	4	3
5 AEMULIS	1	4
6 AGITUR	1	6
7 AGNUM	2	5
8 AGNUS	1	5
9 AMBIGITUR	1	7
10 ANGELICUS	1	8
11 ANGELORUM	1	9
12 ANIMOSA	1	10
13 ANTIQUUM	1	11
14 ASSUMITUR	1	12
15 AUDE	1	13
16 AUXILIUM	1	14

A. Alphabetical listing of words as they appeared in the text. (Note that a serial number precedes the word as listed. First number after each word indicates frequency of appearance in text and the second number refers to the "Main word listing.")

RATIONARIUM VERBORUM

Post singula lemmata proprio numerata numero, vocabula singula numerus praecedei quo in Laterculo primo continebantur, numerus vero subsequetur frequentias singularis ac tandem collectivae.

1 A AB		
1 A	2	
2 AB	1	
2 ACCIPIO ACCIPIS ACCIPERE		
3 ACCIPITE	1	
3 AD		
4 AD	4	
4 AEMULUS AEMULA AEMULUM		
5 AEMULIS	1	
5 AGNUS AGNI		
7 AGNUM	2	

C. Word index combining all A-words with corresponding B-words indicating—to right—frequency of occurrence. (Numbers to left refer to listing B.)

CONCORDANTIAE

1 A AB		
A SUMENTE NON CONCISUS	L 15 43	A
PROCEDENTI AB UTROQUE	P 6 35	AB
IN MORTEM A DISCIPULO	V 2 5	A
2 ACCIPIO ACCIPIS ACCIPERE		
DICENS ACCIPITE QUOD TRADO		
VASCULUM	8 4 15	ACCIPITE
3 AD		
AD FIRMANDUM COR SINCERUM	P 4 23	AD
AD LUCEM QUAM INHABITAS	S 7 28	AD
AD OPUS SUUM EXIENS	V 1 3	AD
VENIT AD VITAE VESPERAM	V 1 4	AD

CONSPECTUS LEMMATUM RATIONARI

1 A AB
2 ACCIPIO ACCIPIS ACCIPERE
3 AD
4 AEMULUS AEMULA AEMULUM
5 AGNUS AGNI
6 AGO AGIS AGERE
7 AMBIGO AMBIGIS AMBIGERE
8 ANGELICUS ANGELICA ANGELICUM
9 ANGELUS ANGELI
10 ANIMOSUS ANIMOSA ANIMOSUM
11 ANTIQUUS ANTIQUA ANTIQUUM
12 ASSUMO ASSUMIS ASSUMERE
13 AUDEO AUDES AUDERE
14 AUXILIUM AUXILII
15 AZYMUS AZYMA AZYUM
16 BELLUM BELLI
17 BENEDICTIO BENEDICTIONIS
18 BIBO BIBIS BIBERE

B. Main word listing. (Note sequential number of these main words and citation of various forms.)

INDEX VERBORUM

1 A AB			
A	L	15	43
A	V	2	5
AB	P	6	35
2 ACCIPIO ACCIPIS ACCIPERE			
ACCIPITE	S	4	15
3 AD			
AD	P	4	23
AD	S	7	28
AD	V	1	3
AD	V	1	4
4 AEMULUS AEMULA AEMULUM			
AEMULIS	V	2	6
5 AGNUS AGNI			
AGNUM	S	2	6
AGNUM	S	3	9
AGNUS	S	21	66
6 AGO AGIS AGERE			
AGITUR	L	6	16

D. Word index combining word entries from A and B with citation of all occurrences in the text. (Numbers to left refer to listing B.)

E. Concordance. (Listing of main words in textual context with identification of location in text. Numbers to left refer to listing B.)

Figure 16-4. Concordance and other listings.

alphabetic list of all of the different forms under which the words used are presented in this text, indicating their frequency. This *laterculum formarum* may be obtained immediately after the words have been alphabetized and arranged. But if the code number of the main word is required, it would be necessary to wait until after the *rationarium verborum* has been prepared. (c) The *rationarium verborum* or *formarum* would be the diagram, systematized and with frequencies, of all the same words regrouped according to their meaning, or more exactly, according to the identity of their functional elements. Such a list is the basis of the author's vocabulary. (d) It would be very simple to list an abridged *conspectus lemmatum*. (e) The *index verborum* will be the index of all the words, or rather of all the word cards, with the reference and arranged according to the *rationarium*. (f) The Concordance will be the same list with the single words followed by the notation as well as the reference. The context may be of one line only, whatever is deemed to be sufficient; but it may also consist of three or more lines; in this case the word in question will always be found in the middle line.

For a simple Concordance, preceded by the *laterculum* and *rationarium formarum*, the required time will, for the first phase, equal the time of one or two typings of the entire text; for the following phases, it will be possible to fulfill in one year that which would take 30 to 40 years of work with the old method. This is the case for the printed Concordance. When, however, one needs to compose a catalog in which the words follow a context of 12 lines on single word cards, 20 or 30 years work can be completed in one year.

In respect to the cost of the work, this much was made clear, on the basis of Italian prices rather than those in the United States: We compared on the one hand a form, composed correctly and paginated, ready to be put into the rotary press, and the on other hand a Concordance obtained from the IBM accounting machine on mats adapted for lithographing, and ready to be passed through the offset system or any other system of lithographic reproduction. We did not include the cost of materials, paper, or zinc. A line of the Concordance prepared and tabulated with the IBM system costs about half what it would cost to set up a line with a linotype or monotype. The computation was made on the supposition that all the work is done in the IBM offices at commercial prices. The difference in the cost will be more appreciable, if one realizes that the cost of conventional printing does not include the cost of preparing the Concordance; while the cost of the IBM listing also comprises all of the work and materials of preparation, such as punching, sorting, and reproducing the cards. In other words, the new method, at half the price required for the preparation of the printing of a Concordance, gives not only the matrices for printing, but also the entire catalog in a flexible form always ready for new studies.

Research on the Structural Elements of Words

A statistical list of prefixes is already contained in the first *laterculum verborum*. It is also very easy to obtain a list of only the first three or four or more letters of words, with totals of frequencies of the single different beginnings of words. It is evident, however, that in this case the machine will list also the short words composed of less than five letters, unless the scholar prevents this by appropriate instructions to the machine.

It is possible to sort from the sets of word cards and form cards those words with particular combinations of initial four or five letters. One can also use the collator in which a pilot card punched with only those letters that constitute the prefixes desired, instructs the machine to extract those cards that contain that composition of initial letters.

If cards are placed in the accounting machine to obtain the list of word cards or form cards, a summary punch can be coupled to obtain a series of cards that represent the various beginnings of words, accompanied by a code number (a serial number representing the alphabetic order) and by the total of the frequencies.

In like manner one can turn to the analysis of the endings. For this we need to punch the words so that the last letter of each word is in the same column. This can be done with the sorter, by separating the words by length, then reproducing all of the cards so that the last letter of each word will be found in the same column. This task is simplified by working from form cards. The words so punched are now alphabetized backwards. This is done by sorting first the initial letter of the longest word and then proceeding from left to right. (This is the reverse of the usual IBM alphabetizing procedure.) We now have the reverse index, in effect a rhyming dictionary. In this way we can list different endings of words indicating the number of frequencies. Also we can pair the accounting machine and the summary punch to obtain a series of cards, which contain the endings of the words, in order to work only on these.

The calculation of the number of letters of a text or of a vocabulary becomes a very simple operation when those words are already punched in cards. One can, for example, use the text cards and explore every single column to sort the letters present in that column. Also we can count each package with the card counter of the sorter, and write the sum totals. To the total of the first column we add the sum of the letters present in the second column and so on. This operation is facilitated by using a sorter provided with a counter, and even more by using the 101 statistical machine.

More work is necessary, but it is still extremely fast compared to manual labor, when one undertakes to analyze the distribution of letters, diphthongs, for example, in words. Such an inquiry, in fact, coincides with the

search for roots of words. Such inquiries can be done by means of successive sortings. This could be shorter if the collator were used, with a pilot card. It would be shorter yet if the 101 statistical machine were used: this machine searches at the same time four different groups of three letters in each word of twelve letters punched in a card. The cards pass through the machine at a rate of 46,000 per hour. The machine does not select the roots only, but rather any combinations of letters specified.

The machine will also facilitate the preparation of materials for study of the distribution of tonic accents, of the proportion of use of the parts of speech and other things. For example: what per cent of nouns, of verbs, of adjectives . . . or what is the predominant structure of the phrase: subject-verb-complement (or predicate), or instead, predicate (or complement)-verb-subject.

Mechanical Search of Phrases

The actual system as described permits searching for a particular phrase, if as emphasized above, the first letter of the preceding word and that of the following word were punched on the word cards. There is, for example, the saying, "sotto questo punto di vista." Among all the cards that carry the word *questo*, the sorter separates those in which *questo* is preceded by *s* and followed by *p*. Among the words *di*, those are selected that are preceded by *p*, preceded in turn by *q* and *s*, and followed by *v*. These cards give the references to all the passages in which are found said sentence, even if the sentence is distributed on two successive cards.

It must be noted here that all that has been said is not necessarily limited to Latin characters. The machines can be provided with any alphabet, and also for Arabic and Hebrew which proceed from right to left. Any series of symbols, signs, or ciphers may be applied to the machines.

The analysis requirements for most texts necessitates the use of punctuation and diacritical marks. The IBM accounting machines such as the 402, 403 and 421 can be utilized for these marks by substituting for the Arabic numerals the desired symbols. This is also the case where card interpretation is required on the IBM 552. Card punches may be modified by substitution of suitable key tops.

For those symbols that should accompany the word, the 12 and 0 zone punch positions of the card should be reserved for use with all IBM accounting machines except the 407. For example, the apostrophe for the articles with elisions in Italian and for the genitive in English, and the point (or period) for the abbreviated words must accompany the word even when it is isolated in the word card. Thus, the German *das ist* becomes abbreviated d. i., but is punched like d-i-; and is also listed as d-i-.

Adoption of the diacritical symbols and of interpretations offers major possibilities through the use of the IBM 26 card punch and 407 accounting

machine, where in addition to the numbers and the letters there are spaces for 11 special characters. Ordinarily such characters are accounting characters, but they can with moderate cost be substituted by signs used in linguistic or philologic studies.

Problems involving accents are more difficult to resolve by ordinary means. Some languages actually present a considerable problem: one thinks of modern French and classic Greek. When special characters for symbols of punctuation, diacritical marks and accents are required at the same time, the top space of a column is not sufficient, except for certain kinds of work and for some languages, for example English, Italian, or Latin. It is necessary then to consider IBM machines which use codes of punches consisting of combinations of two or more columns.

The Near Future

An application of punched cards that should be explored more fully is the automatic tracing of the variants of the same text. The first step in any critical analysis consists of comparing the results of the same analysis applied to a representative selection of various manuscripts or editions. The first line is written down, then the variants as encountered in the other copies. From this material the researcher determines whether the first word is an authentic word of the author. It would be possible to devise a process as follows: Punch line after line for all the editions judged to be representative, one line per card with the reference and proper symbol for each edition. The sorter will assemble all the first lines according to the symbols for the editions, then all the second lines, then the third lines, and so on. The cards ordered in this way are fed to the accounting machine, set to print the group of first lines, then to leave a space before printing the group of second lines, etc. The machine can be so set that for every group, it prints the first line in its entirety, and only those parts of the following lines that are different from each preceding line. Probably it will also be possible for the machine to print, for all of the successive versions of the

```

1 POCA FAVILLA GRAN FIAMMA SECONDA
2 POCA FAVILLA GRAN FIAMMA SECONDA
3 POCA FAVELLA GRAN FIAMMA SECONDA
4 POCA FAVILLA GRAN FIAMMA SECONDA
5 POCA FAVILLA GRAN FIAMMA ASSECONDA
6 POCA FAVILLA GRAN FIAMMA SECONDA
7 PRIMA FAVILLA GRAN FIAMMA SECONDA
8 POCA FAVILLA GRAN FIAMMA SECONDA
9 POCA FAVELLA GRAN FIAMMA SECONDA

```

Figure 16-5. Tabulations of a Set of Hypothetical Variants of a Verse from Dante (Paradiso I, 34)

same line, only that which is different from the first line. Among the graphs is a model showing the tabulation for punching of a set of hypothetical variants of a verse from Dante. (Figure 16-5)

Such a process presents problems of cost. It must be established whether the cost of punching and verifying the same text as many times as there are variations is compensated by the speed and certainty of the subsequent analysis and by the fact that the operation produces text cards for linguistic analysis and the preparation of a concordance of the text.

There are technical problems yet unsolved. The principal problem has to do with prose writings. Due to the fact that the machine exercises control on each column of the card, should a single letter be left out of one line as punched on one card, then the remainder of the text on other cards would be displaced by one column and would as a result be printed as a variant. Such difficulty does not exist for writings in verse, each line of which is started on a new card. For prose works the problem might be resolved by using punched tape or the electronic computer with sufficient memory capacity.

Another development is the printing of the differential context on the back of the word cards. If on all the cards for the words in a given section of text there is recorded the same context, then the first word does not have any preceding context and the last word does not have any following context. As the system is developed there is a need for the context to be printed in such a way that the word punched on the card will be found in the center line of those printed. Thus for all of the words of line 20, the text would begin on line 14 and end on line 25; for the words of line 21, on line 15 to line 26; for the words of line 22, from line 16 to line 27, and so on. This problem is not exactly one of machines for punching cards, but rather one of duplicating machines. However, it is desirable that the processes be linked with the punching on the same cards.

In addition, two parts of the process need to be accelerated: reproduction from the text cards to the word cards, and alphabetic sorting. These two phases notably affect the time and cost of the analysis procedure. Difficult problems of these types obviously do not occur in the ordinary industrial and statistical applications of the IBM system. It is also obvious that, for this reason, an answer to such a demand will come in the future, when research work on linguistics has justified the cost of constructing new models of machines or at least adapting models already in use.

The possibility of punching text cards automatically, starting with the examination of the text by means of a photocell or by other means exists but as yet there are no practical methods for carrying it out. Naturally it will be gratifying when such techniques are operational.

For preparation of concordances, those most valuable kinds of philologic studies, perfection of the method will be achieved when it is possible to

have cards that carry three lines of context punched on the same card. The amount of context carried in a line of 80 columns is not sufficient in a printed concordance prepared automatically. For the most part it is necessary to have a context of three lines, so that the word in question is always in the middle line. As already described, it is possible to obtain such print-



Words are inverted through
Control Panel wiring

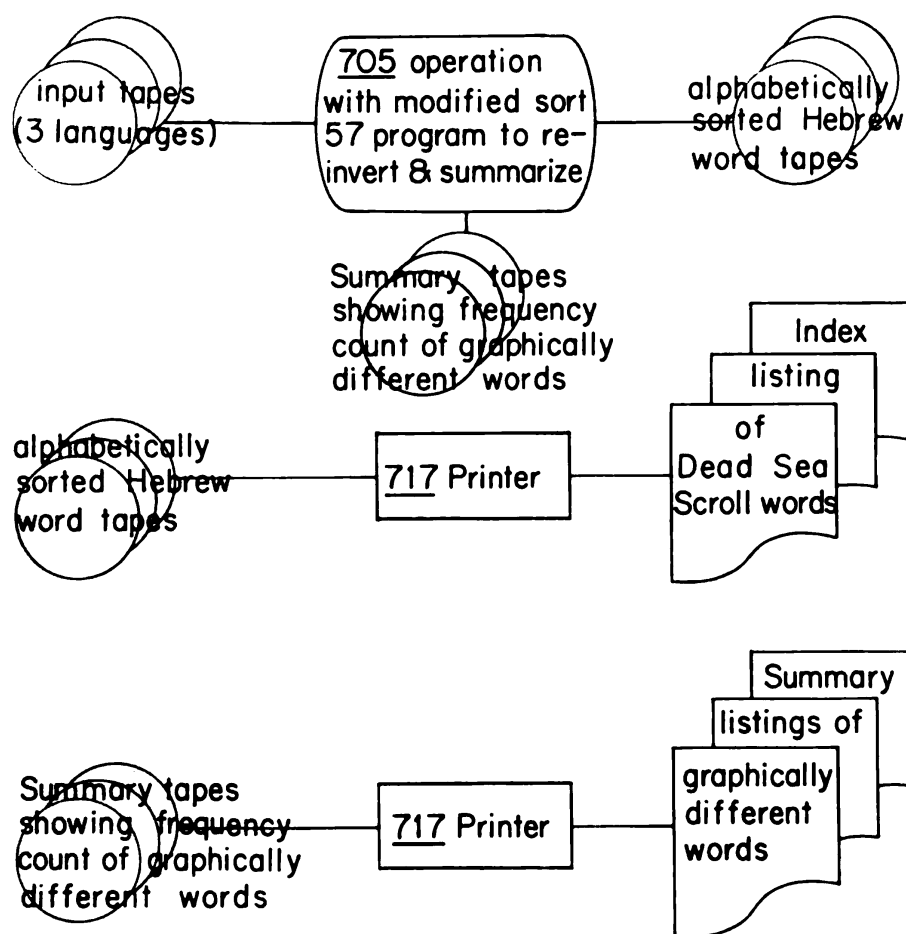


Figure 16-6. Simplified block diagram of "Dead Sea Scrolls" processing on EDPM equipment.

ing even with standard machines, but repetitive use of the sorter and collator is necessary as well as extensive operations with the accounting machine. When such extensive operations can be avoided by using cards containing three lines of context, or some equivalent means, the mechanization of linguistic analysis can be said to have reached that stage where substantial change will not be required for some time.

When punched card systems operate so that the cards are passed through the machines along the cards' long axis, which is perpendicular to the axis of motion through present day machines with exception of the punchers and verifiers, then it would be possible to work on the basis of successive circuits, like a telephone center. The demands of linguistic analysis would then be satisfied even more completely, faster, and more economically.

Such observations on desired developments should not overlook the fact that even with standard machines the punched card system permits more extensive, more certain, more advanced and more economical studies than would have been possible except with the patient work of many men.

For similar work in the preparation of concordances by machine, the reader is



Figure 16-7. Father Roberto Busa comparing the words of a modern scribe—the printing unit of an IBM 705 computer—with those written two thousand years ago by scribes of an ancient Hebrew sect living near the Dead Sea.

referred to the work of Reverend James Ellison in preparing a concordance of the Bible by means of a computer. See, for example,

"According to Mark 4.—," *Time*, vol. 64, August 9, 1954, pp. 68-9

Soule, G. "Machine That Indexed the Bible," *Popular Science*, vol. 169, November 1956, pp. 173-5

"Bible Labor of Years is Done in 400 Hours," *Life*, vol. 42, February 18, 1957, p. 92

Cook, C. M. "Automation Comes to the Bible," *Christian Century*, vol. 74, July 24, 1957, pp. 892-4

Appendix

Work is now in process in Gallarate and in New York in applying the method of literary analysis described here to the task of cataloging the "Dead Sea Scrolls." The nearly thirty thousand words under study were punched into IBM cards. A card was punched for each word, indicating its exact location and distinguishing characteristics. The entire set of cards was converted to two reels of magnetic tape by the IBM 705 computer in approximately two hours. See Figures 16-6 and 16-7.

In order for the IBM 705 to perform the indexing of the "Dead Sea Scrolls," the following items had to be taken into consideration.

1. Card input requirements.
2. Printed output requirements.

Since Hebrew words are right-most justified, read and interpreted, special considerations had to be dealt with prior to obtaining the desired results. For the input, the Hebrew word cards are initially converted to tape in such a fashion that the words will be recorded on the magnetic tape in an inverted form (left-most justified).

The Hebrew words range from 1 to 12 character positions.

The magnetic tapes, once created, are then loaded on the IBM 705 and with the aid of modified Sort 57 Program Deck, the following has been accomplished.

1. This program first sorts all these Hebrew words alphabetically and at the same time re-inverts them into their original form prior to writing them on the output tape.

2. An extension to the program provides for creating a summary word tape on which are written only those words which are graphically different from each other. This summary record will also show an identification serial number with the frequency of usage of each word.

Later, on an off-line basis, these tapes will be listed on a tape-to-printer operation. The total off-line printing time is five hours.

Chapter 17

AN ABSTRACTING AND INFORMATION SERVICE FOR PLANT BREEDING AND GENETICS

R. H. RICHENS

*Commonwealth Bureau of Plant Breeding and Genetics, School of Agriculture
Cambridge, England*

A coding system in which all published articles on plant breeding and genetics are coded on punched cards for future reference is in use at the Commonwealth Bureau of Plant Breeding and Genetics, Cambridge. Since the principles involved appear to be of general interest for scientific study, an account of the technique used is being given here, in the hope that it may prove useful to others working in this field.

General Principles of Mechanized Coding

Mechanized coding of scientific papers involves the following processes: (1) Construction of a coding dictionary giving the equivalent in the code of any feature of a scientific paper. (2) Selection of those features of the paper that are to be coded. (3) Assigning to each of the chosen features its equivalent in the code. (4) Rendering the code into a medium susceptible of mechanized manipulation. It is necessary to consider each of these processes in further detail.

The construction of a coding dictionary must obviously precede the setting up of a mechanized coding system. A coding dictionary consists essentially of a series of entries giving for each possible feature of a scientific paper a unique equivalent in the code. The number of ways in which this might be accomplished is obviously very large.

The simplest system would be to reproduce the feature to be coded without alteration. Thus, it might be agreed to code the word "wheat" when it occurs in a scientific paper by reproducing the word "wheat" in the code. Although this method has the advantage of extreme simplicity, subsequent mechanized manipulation is extremely intricate. The procedure could be refined by translating all linguistic equivalents of "wheat," such as "blé," "Weizen," "trigo," "grano," or "pszenica," into one particular language—English, and it would be possible to go some way toward eliminating synonymy within a language, or within a scientific jargon. It could be decided, for instance, that "*Triticum vulgare*" should be used instead of "wheat," or vice versa.

Brevity is advantageous in most coding systems, especially when these are to be mechanized. Therefore, if it is possible to replace an expression, such as "*Triticum vulgare*," by a unique equivalent with fewer letters, the efficiency of the coding system would probably be increased. There is no reason why the coding equivalent should be a succession of letters at all. It could be replaced by a series of numerals, a mixture of numerals and letters, a Chinese character, or some hieroglyph with no previously accepted significance. Convenience, however, and ease of recognition demand that the symbols used should be familiar, and it is, therefore, hardly necessary to consider coding equivalents made up otherwise than by the juxtaposition of letters and numerals.

So far consideration has been given to a very simple type of coding, that in which single words occurring in a paper are coded. In most cases, however, it is not the vocabulary of a paper but the ideas expressed that are coded. It is obvious that an idea cannot be reproduced as such in a code since ideas are characteristics of minds. All that can be done is to devise a code giving unique equivalents to each idea; these may be words or symbols as mentioned above.

If ideas are being considered, a further very important refinement can be introduced into the coding system. Ideas exhibit manifold logical relations, one with the other. These relations, or some of them, can be reproduced in the structure of the coding equivalents. Thus, the idea of plant includes the notion of wheat, or, in logical terminology, wheat is a subclass of plants. Everything that is a member of the class of wheats is a member of the class of plants. Class inclusion occurs prominently in a system such as the Universal Decimal Classification (U.D.C.), in which the coding equivalent for wheat, 633.11, is so constructed that it indicates that the wheat class 633.11 is a subclass of cereals 633.1, and the cereal class, in turn, is a subclass of crop plants 633.

The U.D.C. is only one, and, from a logical point of view, a somewhat unsatisfactory coding system. Alternative and better systems could be devised. Not only could additional logical relations between concepts be indicated in the structure of the coding equivalents, but a more complex and comprehensive system of class-inclusive relations could be used. The particular merits and demerits of the U.D.C. will be considered in the following.

The second process in mechanized coding is the selection of those features of the original paper that are to be coded.

The simplest case would again be constituted by vocabulary analysis. By selecting all the most scientifically significant words in a paper, a fairly adequate and detailed indication would be given of its contents. Moreover, this operation is in itself almost entirely mechanical, and requires no more than an ability to pick out significant words. If the procedure were reversed

so that all the words in a scientific paper in common speech—its “basic” vocabulary—were eliminated, and the residue taken as the significant vocabulary, the whole coding operation could be performed by a suitably constructed machine.

To take an example, the scientifically significant vocabulary of a paper by Arnason, Cumming, and Spinks is as follows¹

aberrant	microcurie
aberration	<i>monococcum</i>
acetocarmine	mutation
<i>aestivum</i>	Pelissier
anaphase	phosphate
anther	phosphorus
beta	pollen
chromosome	radioactive
diaphane	radium
<i>distichon</i>	spikelet
<i>durum</i>	telophase
einkorn	tetraploid
gamma	Thatcher
Hannchen	translocation
hexaploid	<i>Triticum</i>
<i>Hordeum</i>	<i>vulgare</i>
inversion	X-ray
localization	

These words indicate quite clearly the nature of the contents of the original paper, which is entitled “Chromosome Breakage in Plants Induced by Radioactive Phosphorus (P³²)”.

Usually, however, the aspects of a paper to be coded are decided by a reader able to understand what the paper is about. The coder classifies the paper under certain general heads. He fits the paper into its place in a general classification of knowledge, however ill-defined the latter may be. For instance, in the case of the paper by Arnason, Cumming, and Spinks this could be classified under three general heads: anomalous nuclear changes, botanical effects brought about by chemical agents, and radioactivity. In this classification three broad classes of phenomena have been selected, of which the principal phenomena described in the paper are representative members.

It may seem that coding by vocabulary analysis and coding by meaning differ fundamentally, but it is possible to exaggerate this difference. In theory, at least, it would be possible to devise a mechanical procedure based on an analysis of the vocabulary and syntax of scientific papers, in which coding equivalents corresponding to the general classificatory heads just mentioned were assigned to the paper whenever certain combinations of

¹ Arnason, Cumming, and Spinks, *Science*, 107, 198–99.

words occurred, or even when certain words appeared in certain syntactical relations. Thus, it might be arranged that a paper containing any three of the following words, aberrant, aberration, deletion, interchange, inversion, nucleus, and translocation, should be given a coding equivalent corresponding to the general head "anomalous nuclear changes", and similarly for other groups of words. This means that a purely mechanical analysis of vocabulary might be devised which would have the same result as a classification under general heads assigned by an understanding reader.

Attention has been given the seemingly impractical method of vocabulary analysis for coding scientific papers because it has theoretical interest and because it shows that it is possible to obtain mechanically results normally achieved only by making use of an understanding reader. This point may be put more precisely by stating that vocabulary analysis and semantic analysis of a scientific paper may produce isomorphic classifications. As far as practical problems go, however, the extraordinary efficiency of semantic analysis performed by the mind, as compared with the cumbrousness of purely mechanical methods, suggests that the latter are unlikely to be of practical importance in any small-scale scientific information center.

Having now considered both the construction of coding dictionaries and possible ways of selecting the features of a paper that might be usefully coded, there is little to say on the third coding operation mentioned above, namely, assigning to each of the selected features of a paper its equivalent in the code. It is possible, since assigning coding equivalents is the setting down of a series of one-one equivalences, to construct a machine to do it; but here again a mechanical model is only likely to prove more efficient than the mind when the scale of the operation becomes very large. For comparatively small-scale projects a human operator, using a check list and his memory, is likely to provide the most efficient means of assigning coding equivalents to the selected features of a scientific paper.

The last process in mechanized coding, rendering the code into a medium susceptible to mechanical manipulation, has many possibilities. It is in some ways unfortunate that handwriting and print cannot be manipulated mechanically without conversion, or translation in the logical sense, into a different medium. It is probable that the irregularities in handwriting and type setting are too great to enable any but the most complicated mechanical models to deal with their contents as efficiently as a human operator. It therefore seems necessary at the moment to render codes into appropriate functional media by employing a human intermediary.

It is possible that the same individual may select the features of a paper to be coded, assign the coding equivalents, and render the latter into a functional medium. This is not necessary, however, and each of the processes could be performed by a distinct individual. Only considerations of effi-

ciency can decide how the human material in a coding system is to be disposed.

Coding media present many possibilities, only a few of which have immediate practical interest. These are punched cards, photo-electrically detected light signals and electromagnetically sensed magnetic signals. The two latter media require apparatus beyond the reach of a small installation. They offer immense possibilities for very large-scale scientific information centers, but for small installations punched cards alone seem to offer an economical method. In this chapter, which is only concerned with coding methods suitable for small installations, attention will be confined to punched-card techniques.

*Outline of the System under Trial at the Commonwealth Bureau of
Plant Breeding and Genetics*

- (1) The coding dictionary used is the U.D.C.
- (2) Selection of the features of scientific papers to be coded is undertaken by the scientific staff of the Bureau.
- (3) The person who decides what aspects of a paper should be coded also assigns the coding equivalents, that is, a series of U.D.C. numbers.
- (4) The U.D.C. numbers are punched into cards by the clerical staff of the Bureau.

Merits and Demerits of the U.D.C. for Mechanized Coding

It has already been mentioned that the U.D.C. is a noteworthy example of a coding system whose structure is logically significant. There are several considerable advantages attached to its use. These are as follows:

- (1) The U.D.C. is very comprehensive, and a large proportion of the features of scientific papers that will require coding are to be found in it. Most organizations dealing with scientific information are interested primarily in the scientific aspect of their work, and not in logical principles of classification. For them, a ready-made classification, even though unsatisfactory on several scores, is likely to be of great service.
- (2) The U.D.C. is extensible, and its subject headings may be subdivided indefinitely.
- (3) The U.D.C. is internationally recognized.
- (4) The way in which U.D.C. numbers are constructed lends itself well to punched-card techniques.

Against these advantages, there are three disadvantages:

- (1) It is not possible to draw together subjects which are placed far from each other in the U.D.C., and to subdivide the logical sum of the two classes, cytology + genetics for example, as a single unit. This is a particular disadvantage in regard to the marked syncretic tendency of scientific studies. Subjects which once appeared remote from each other, as cytology, genetics, virus research, cancer research, the theory of evolution, paleontology, and ethnobiology, have now converged to such a degree that a discovery in any

particular one is quite likely to be of significance to one or more of the others.

(2) The same subject is liable to appear in more than one place in the U.D.C. schedules. Thus, heredity is coded as 575.1 and 581.169, while hybridization is entered under 575.12 and 631.523. For purposes of mechanized coding, those double entries have to be expurgated, and a choice made as to which alternative number in each case is to be used.

(3) In some cases in the U.D.C. the principle of indicating the subclasses of a subject heading by suffixing further digits to it is not adopted. Thus, although basidiomycetes are given the number 632.44, the hemibasidiomycetes and eubasidiomycetes receive the number 632.45 and 632.47, respectively, instead of suffixing digits to 632.44.

Apparatus

"Powers-Samas" cards with 65 columns are used. Ten different colors are employed. The cards are punched by means of an ordinary "Powers-Samas" hand punch. The remaining item of equipment is a "Powers-Samas" sorter, having a selective sorting attachment.

Coding Procedure

In the Commonwealth Bureau of Plant Breeding and Genetics all obtainable scientific publications of possible interest to plant breeders and geneticists are read by the members of its scientific staff, who prepare for each paper an abstract in English for the Bureau publication, *Plant Breeding Abstracts*. In addition to summarizing each paper the abstracter also gives to it a U.D.C. code number as a guide to the nature of its contents. Papers that are only of indirect interest to plant breeder and geneticists are coded but not abstracted.

The method of using the U.D.C. for coding purposes is the customary procedure of assigning a series of decimal numbers, each of which corresponds to one aspect of the paper being coded. Thus, referring back to the paper of Arnason, Cumming, and Spinks, already mentioned, this might be given the U.D.C. decimal numbers:

576.356	} corresponding respectively to the three general heads	{ anomalous nuclear changes botanical effects brought about by chemical agents radioactivity
581.04		
539.16		

The coding numbers assigned by the Bureau are not exhaustive or even general descriptions of the contents of the papers coded. Papers are classified from a particular point of view, that of its readers, and it is possible that their interest might be confined to a chance reference to plant breeding

in a paper covering principally some other topic. In this case, the U.D.C. numbers assigned by the Bureau will refer only to the relevant paragraph. A scientific information organization dealing with scientific papers in general or from a different point of view would code such a paper quite differently.

Layout of Card

It has been found by experience that the subject matter of papers on plant breeding and genetics can be satisfactorily coded in most cases by one or more decimal numbers beginning with the following two digit combinations:

51	mathematics
53	physics
57	biology
58	botany
63	agriculture

In the layout of the card each of these subjects is assigned a separate panel of five columns (see Figure 17-1a), while for the last item, agriculture, under which several subordinate aspects require separation, three five-column panels are used:

631	agronomy
632	phytopathology
633-5	crops

Should any U.D.C. numbers other than the above need to be employed, they are allocated to a separate five-column panel, entitled "General". There are also certain suffixing decimal numbers which may be appended to any of the principal decimal numbers, such as .01 for bibliography. Such numbers are allocated to a further five-column panel, entitled "Tags".

All the panels of the card concerned with subject matter are of the same width—five columns. As far as the requirements of the coding scheme are concerned, this number of columns per panel is unnecessarily large in some cases. For instance, for the purposes of coding papers on genetics and plant breeding, the 51 and 53 panels dealing respectively with mathematics and physics could have been reduced. It has been found, however, that the work of punching is rendered easier if each of the subject panels is of equal width. In determining the precise layout of the card psychological considerations must not be disregarded.

The remainder of the card is occupied by a three column panel for representing the decimal number of the country to which a paper refers, a three-column panel for representing its year of publication, a two-column panel for indicating the number of the volume of *Plant Breeding Abstracts*, in

which a summary of the paper has appeared, and a final one-column panel, entitled "Class," in which a hole is punched in position 1 to indicate that the card is a subject card, thereby differentiating it from the other series of punched cards maintained by the Bureau. In the case of the three-column panel for the year only the three final digits of the year, for example 948 for 1948, are punched. This will suffice for the next thousand years. Subsequent years can be differentiated by punching additional holes in the positions above 0.

It is important to realize that the layout of the card is one of the most weighty factors determining the successful application of punched-card methods to scientific information work. A single mistake in layout will prejudice the whole of subsequent work. In general, an organization dealing with a comparatively narrow and highly specialized field will need a small number of wide panels. Those dealing with a wider range of subjects should use a larger number of narrower panels. Care should be taken that the frequency with which two decimal numbers are liable to fall within the same panel is low. It is also advantageous to have the panels of equal widths, even if it wastes space, since, as already mentioned, this lightens the work of the punching clerks.

Preparation of Punched Cards

After a paper has been coded by the scientific staff of the Bureau, the U.D.C. numbers are sent in to the clerical staff, together with the original paper to which they refer. The punching clerk then punches each of the digits of the U.D.C. numbers into the appropriate panels of the punched card by means of a hand punch, punching also the year of publication of

The diagram illustrates a rectangular punched card with a header section at the top. The header is divided into two main parts: "ONLY FIRST TWO DECIMALS" on the left and "PUNCH IN FULL" on the right. Under "ONLY FIRST TWO DECIMALS", there are six panels labeled 51, 53, 57, 58, 631, and 632. The panel 632 is followed by a panel labeled 633- with a vertical line to its right. Under "PUNCH IN FULL", there are four panels labeled GENERAL, YEAR, MONTH, and DAY. Below the header, the card is divided into a large central area with a grid of small circles, representing punch positions. The grid is organized into rows and columns, with some columns containing multiple circles. The card is framed by a border with small circles along the edges.

Figure 17-1(a)

Figure 17-1. Stages in the preparation of subject and author cards. *a.* Subject card before punching. *b.* Card after punching. *c.* Card after typing. *d.* An author card.

Figure 17-1(b)

ONLY FIRST TWO DIGITS						SEARCH IN FULL					
51	53	57	58	631	632	633-	CARD	PAGE	SHEET	VOL.	FILE

(L) ARKASON, T.J.,
CUDWING, E. and
SPINKS, J.W.T.

Chromosome breakage in plants induced by radioactive phosphorus (³²P). Science 1948; 107: 198-99.

XVIII(4) Summary. R.

Figure 17-1(c)

ARNASON, T. J., CUMMINGS, E. and SPINKS, J. W. T.	576.346:591.04:539.16
(L)	Chromosome breakage in plants induced by radioactive phosphorus (P^{32}). Science 1948; 107: 198-99
XVIII (4)	Summary. R.

Commonwealth Agricultural Bureau

Figure 17-1(d)

the paper, the volume number of "Plant Breeding Abstracts," in which the paper is noticed, and a hole in position 1 of the "Class" column to indicate that the card is a subject card. Figure 17-1b shows the mode of punching the coding data of the paper by Arnason, Cumming, and Spinks mentioned

previously. The first two digits of decimals coming in any of the first seven panels are not punched. The heading of the panel in these cases provides sufficient indication.

After punching, the subject card is put in a typewriter above a standard filing card, with a carbon in between. The following information is then typed in duplicate on both cards:

- (1) The name of the author or authors.
- (2) The title of the paper, together with a translation if the original is not in English.
- (3) The other bibliographical details of the paper.
- (4) The volume of "Plant Breeding Abstracts," in which the article is noticed.
- (5) The initials of the abstracter.
- (6) A note as to the library in which the original paper is deposited.
- (7) The coding decimals of the paper.

The preparation of the subject card is now complete (Figure 17-1c). The standard filing card (Figure 17-1d) is used as an author card.

The subject cards used are of ten different colors for convenience in filing. The color of the card is determined by the principal decimal number of the card, that is, the number recording the particular crop being treated. If there is no entry in the 633-5 crop panel in the coding details of the paper, that is, if the paper is a general one not dealing with any specific crop plant, uncolored Manila cards are used. In all other cases the color is determined by the final digit of the crop decimal, according to the following scheme:

Final Digit	Color
1	Red
2	Orange
3	Yellow
4	Green
5	Blue
6	Mauve
7	Grey
8	Pink
9	Brown

Thus, any paper on wheat, which will receive as one of its coding decimals, the number 633.11, will be entered on a red card since the final digit of 633.11 is 1. Similarly, papers on potatoes, 633.491, will also receive red cards, while papers on barley, 633.16, will be given a mauve card.

Should a publication deal with a number of unconnected subjects, as for example, the annual report of an agricultural research station, a series of cards will be made, one for each of the subjects included. Sometimes the coding data of a paper on a single topic contain two or more decimals that would come under the same panel, an eventuality which the card is designed to avoid as far as possible. When this happens, one decimal is punched in its proper panel, and the second (together with the first two digits normally

unpunched) is punched in the "General" panel. A second card is then punched in which these two numbers are reversed. Thus, a paper by Wellensiek on "Methods for Producing *Triticales*"² was assigned the five decimal numbers:

633.11	} corresponding respectively to the subject heads	wheat
575.129		true breeding hybrids
633.14		rye
581.04		botanical effects brought about by chemical agents
(49.2)		Holland

For this paper two cards were prepared, the first (Figure 17-2a), a red card with the U.D.C. number 633.11 punched in its normal panel and 633.14 punched in the "General" panel, and a green card (Figure 17-2b) with 633.14 punched in its normal panel and 633.11 punched in the "General" panel. In both cases the suffixing decimal number .04, is punched in the panel headed "Tags," as already explained.

When two or more authors collaborate in writing a paper, a separate author card is made for each.

Filing

The author cards are filed in alphabetical order in the usual way.

The punched subject cards are filed according to crop since experience has shown that plant breeders' inquiries normally relate to specific crops. Since the subject cards are colored according to crop, the cards will also be filed in color order. Within each color group corresponding to one U.D.C. number in the 633-5 crop panel the cards are not kept in predetermined order. Thus, all the subject cards relating to wheat, 633.11, are not filed in any particular order. Since all the cards pertaining to a particular crop are of the same color and the colors representing adjacent crops are distinguished according to a standard plan, all that a filing clerk has to do when inserting a fresh card is to locate the approximate position of the crop, and insert the card anywhere within the run of cards of the same color already filed.

Extraction of Information

Requests for scientific information come in under various forms. Any inquiry relating to the published works of a particular author will be answered from the author file. All other inquiries are dealt with by passing appropriate subject cards through the sorter, and sorting either by a single column or by ten adjacent columns simultaneously by means of the selective-sorting attachment.

² Wellensiek, "Methods for Producing *Triticales*," *Journal of Heredity*, 38, 167-73.

ONLY FIRST TWO DECIMALS						PUNCH IN FULL						
51	53	57	58	631	632	633-	GENERAL	TAX	MONTH	YEAR	VOL.	PAGE
WELLESIEK, S.J.						633.11:575.129:633.14:581.04(49.2)						
(L) Methods for producing <u>Triticale</u> .												
J. Hered. 1947: 36: 167-73.												
XVIII(1) Summary. T.A.												

Figure 17-2(a)

ONLY FIRST TWO DECIMALS						PUNCH IN FULL						
51	53	57	58	631	632	633-	GENERAL	TAX	MONTH	YEAR	VOL.	PAGE
WELLESIEK, S.J.						633.11:575.129:633.14:581.04(49.2)						
(L) Methods for producing <u>Triticale</u> .												
J. Hered. 1947: 36: 167-73.												
XVIII(1)Summary. T.A.												

Figure 17-2(b)

Figure 17-2. Duplicate subject cards of a paper to which two decimal numbers have been assigned which would normally be punched in the same panel. *a.* A red card with 633.11 punched in its normal place and 633.14 punched in the "General" panel. *b.* A green card with these two decimals reversed.

If, for instance, an inquirer should require a classified list of papers on cereal diseases, all the cards bearing the U.D.C. numbers 633.11 to 633.19 would be extracted from the file and passed through the sorter set on column 27, the second column of the 632 phytopathology panel. By this operation eight batches of cards would be segregated, dealing respectively with physiological diseases, galls, bacteria, fungi, angiosperm parasites and weeds, destructive animals other than insects, insects and viruses.

Should any particular category in a one-column sort be of no interest, it can be ignored in the sort by suitable adjustment of the sorter.

More frequently, an inquiry will take a more specific form, and may, for

example, request details of all publications dealing with the resistance of wheat varieties to low temperature. To answer this all the wheat cards which bear the decimal 633.11 would be extracted from the file; their red color makes them easily discernible. The selective-sorting attachment is now fitted on to the sorter and set for the numbers 152162111 on columns 21 to 29, respectively. By passing the wheat cards through the sorter it will now pick out the cards relating to all papers in which resistance to unfavorable conditions is mentioned, these having been given the decimal number (63) 1.521.6 beginning at column 21, the first column of the 631 agronomy panel. However, these cards will be extracted only if they deal at the same time with low temperature, which is recorded by the decimal number (63) 2.111 beginning at column 26, the first column of the 632 phytopathology panel. The panels of the cards are arranged so that subjects likely to occur in combination lie near one another. The extracting of the wheat cards from the file and the single passage through the sorter will now have provided information on all articles dealing with resistance to low temperature in wheat.

Should the inquiry take an even more specific form, for instance, a request for all information on the genetics of resistance to low temperature in wheat, a visual inspection of the cards extracted as above may suffice. Alternatively, a second passage of these cards through the sorter, with the selective-sorting attachment reset for genetics (57) 5.11 on column 11, the first column of the 57 biology panel, will serve the same end. Whenever two passages of the cards are required for selective sorting, the less frequently treated subject should be sorted out first, as the number of residual cards for the second selective sort will then be smaller. This rule applies also to hand-sorted cards, as noted in Chapter 2.

Inquiries relating to papers concerned with particular countries or published at a particular time or over a particular period may be dealt with by setting the selective-sorting attachment over the "Country" and "Year" panels. It is also possible to eliminate unwanted categories during selective sorting. Thus, should an inquirer ask for all the papers dealing with rye that had appeared between 1940 and 1948, but excluding those that concerned North America, a single passage of the rye cards through the sorter, with the selective sorter set for 1940-48, but excluding cards bearing a hole in position 7, the U.D.C. number for North America, would give the required information.

Standing Inquiries

A scientific information unit may be required to supply its clients personally with information of particular interest to them at regular intervals. This can be done mechanically by preparing one or more pilot cards for

each client and punching these with the coding equivalents of the subjects in which interest is expressed. Ever so often, the entire set of pilot cards is meshed with the subject cards acquired since the last similar operation. In this way, each pilot card comes at the head of a run of subject cards carrying the same punching and therefore dealing with the same subject as that referred to on the pilot card. The bibliographical details given in the run of subject cards can then be copied out and sent to the client or clients entered on the pilot card.

Subject Index

The annual subject index of an abstract journal is usually long and arduous to compile. The Bureau has mechanized part of the process. The subject index takes the form of primary heading, subdivisions and cross references. Each coding equivalent has a corresponding index entry, which can be used either as a primary heading or subdivision. It also has, in most cases, a series of cross references attached to it. The subdivisions of the primary headings corresponding to any one panel of the subject card are arranged so that all are in one other single panel or, at least, in no more than three other panels.

In making the index, all the subject cards of the year concerned are extracted and sorted first into numerical order of one of the panels acting as subdivisor, and then into numerical order in the panel of the corresponding primary headings. All the cards relating to one particular set of primary headings and subdivisions are now assembled. A slip can now be made for each primary heading plus subdivision and the references to it can be transferred by hand from the relevant subject cards. This can be expedited in some cases by using accessory punched cards, punched so as to be sortable at the end of the whole operation into alphabetic order. Cross references can be made at this stage.

The operation is then repeated for every pair of panels corresponding to a combination of primary headings and subdivisions. The slips can then be hand-sorted, or if accessory punched cards have been used, these can be machine-sorted, to give the final subject index, complete with primary headings, subdivisions and cross references.

Part III

FUNDAMENTAL CONSIDERATIONS IN CODING AND SYSTEMS DESIGN

Part III

FUNDAMENTAL CONSIDERATIONS IN CODING AND SYSTEMS DESIGN

CHAPTER 18

SUBJECT MATTER ANALYSIS AND CODING—SOME FUNDAMENTAL CONSIDERATIONS*

JAMES W. PERRY

*Center for Documentation and Communication Research
Western Reserve University, Cleveland, Ohio*

Introduction

Valuable information often accumulates in amounts that require us to use aids to human memory. To serve their purpose efficiently, such aids must be organized in an orderly fashion. The development of such aids is a very old problem. The ancient Babylonians had classification systems for their libraries of clay tablets. Alphabetized indexing is at least as old as Gutenberg's invention of printing. In recent years, the application of new tools, e.g., punched cards and electronic machines, have enabled new methods to be developed for establishing and using orderly organization of information. To develop new methods in the different forms that provide optimum advantages in dealing with different situations and circumstances, it is essential that the age-old problem of orderly organization of information shall be investigated anew in the light of the unusual capabilities of the new tools that are now finding increasingly widespread application.

The organization of information for subsequent use is based on one form or another of the analysis of the information as to ideas, concepts, notions, abstractions, relationships, and the like. The procedures for conducting such analysis and, more particularly, the form of recording its results vary widely depending on the tools that we may choose to employ. For example, if we decide to designate various characteristics of the subject contents of documents by words and phrases and arrange them in alphabetical order, we have an index. If our analysis of documents as to characteristics is used to arrange the documents in groups and subgroups according to similarities and differences, we will establish a classification.

These conventional methods of arrangement, indexing and classification, have been developed and used for the orderly arrangement of things in space. These methods function by establishing positional locations, in

* This research was supported in whole or in part by the United States Air Force under Contract No. AF 49(638)-357 monitored by the AF Office of Scientific Research of the Air Research and Development Command.

ordered arrays on sheets of paper, on shelves or in drawers, or in separate pigeon holes. Arrangements in three-dimensional space encounter practical limitations. For example, the classified shelving of books cannot simultaneously group them according to year of publication, language and the many diverse subjects to which they pertain. Alphabetized indexes would be excessively bulky and prohibitively costly if they attempted to list every combination of characteristics which analysis of the subject contents of documents may reveal as important. Punched cards and electronic machines enable us to surmount such limitations of three-dimensional space. These newer tools are essentially multiple dimensional in character by virtue of their ability to search out and to select information on the basis of new combinations of characteristics, that is to say, combinations not formulated or established at the time the information is analyzed. Thanks to their multi-dimensional character of operation, the newer tools, as exemplified by punched cards, enable us to center our attention on *selecting* and *retrieving* information for use. We are thereby released from the limitations of systems which function by deciding *where* we are going to put documents within a fixed array of one type or another.

It is, of course, possible to base the operation of punched cards, electronic selectors and similar devices on the analysis of information in the form of a conventional subject matter classification or a simple index consisting of single terms: But more sophisticated concepts of subject matter analysis are required to take *full advantage* of these mechanical and electrical devices.

We must break out of our mental strait jacket which has been imposed by accustomed means of subject handling: one-dimensional rows of words; two-dimensional sheets of paper; three-dimensional shelves and pigeon holes. We must stop thinking about *where* to put things. With punched cards and machines, we can do the equivalent of putting one thing in several places at the same time, of putting one thing just anywhere and retrieving it at will, of putting several things together somewhere and retrieving something different.

The "things" just mentioned are ideas or concepts that characterize information, the "somethings" are combinations of ideas or concepts and relationships and the "places" may be exemplified by holes and notches in pieces of cardboard. In more general terms, the "places" are meaningfully disposed shapes and discontinuities in pieces of matter, or energy conducting, generating or modifying spots and areas, such as magnetized spots on wire or cylinders, electric current conducting marks, light conducting, reflecting or absorbing spots and shapes, radioactive or fluorescent areas. To make use of such "places," we must develop a code which assigns a unique pattern of discontinuities to designating each idea or concept, i.e., to each

characteristic of subject matter. The construction and manipulation of the array of discontinuities is done by appropriately designed mechanical or electrical devices, described in other chapters. This chapter will treat subject matter analysis and coding for the *selection* of desired information and its concomitant *correlation*.

Statement of Basic Operations

Other chapters in this book, especially Chapter 3, have directed attention to a variety of devices and equipment that have been applied for searching, selecting and correlating recorded information. On the one hand, we have relatively simple devices such as hand-sorted punched cards and, at the other extreme, fully automatic electronic selectors and computers. In spite of obvious diversity in design, the various devices and machines have certain general operational functions in common. An understanding of basic principles underlying such operational functions is essential to developing codes for achieving efficient application of a given device or machine. Such understanding can also provide guidance in selecting appropriate equipment for a given set of requirements and circumstances.

The descriptions and summaries of practical applications as presented in this book make it clear that the use of various types of devices and equipment always involves four basic operations:

1. Analysis of the subject contents of graphic records.
2. Recording the results of analysis in an appropriate searching medium, such as, hand-sorted punched cards, magnetic tape, etc.
3. Analysis of information requirements in terms of operations to be performed.
4. Performance of searching and selecting operations.

The first two steps are preliminary and preparatory in nature. The necessity for performing these steps arises from the limitations of machines that are available at present or that could be constructed at reasonable cost by exploiting present-day technology. Developments in character-recognizing devices may provide, within a few years, machines that are capable of scanning printed material and detecting individual words, punctuation marks, etc. Such detection is, however, quite different from selection of those words and phrases that correspond to important aspects of subject matter, and that might conceivably form the basis for writing abstracts, preparing indexes or encoding for machine searching. Furthermore, particularly with scientific and technical subject matter, diagrams, maps, graphs, equations and formulas constitute and often present a major part of the subject matter and its correlation with textual material for abstracting, indexing, and encoding purposes usually requires a high degree of understanding of the

subject matter.¹ Programming a machine so that it provides such understanding appears quite impossible with available equipment—or equipment that can be anticipated in the foreseeable future. To make the most effective use of the machines with which we can expect to be working during the years ahead, it is necessary to provide an interpretation of the subject contents of graphic records in a form amenable to selecting and correlating operations that are performed by automatic and semi-automatic devices. A system for achieving such interpretation is often referred to as a “code” and its importance warrants considerable further discussion.

The third and fourth steps outlined above pertain to the application of various mechanical or electrical devices to collections of encoded information to identify, to select and to correlate items of pertinent interest. The third step—the analysis of information requirements—is, in character at least, closely akin to the preliminary encoding of input information. Just as searching machines available at present require that input information shall be preliminarily analyzed and encoded, so also presently available devices and equipment require that questions to be answered and requests for information in general shall be submitted to preliminary analysis. Such analysis must be conducted on the same basis as the preliminary analysis and coding of the information to be searched. The results of such analysis of questions and information requests serve to guide the selecting and correlating operations to be performed by various mechanical and electrical devices. With hand-sorted punched cards, for example, the analysis of information requirements will lead to decisions to perform one or more sorting operations directed to various punching positions. When a plurality of such sorting operations are involved, the analysis of information requirements may suggest possibilities for conserving considerable time and effort in manual manipulation of the cards. With more elaborate equipment, particularly fully automatic electronic selectors, the analysis of information requirements provides instructions for conditioning the machine to perform desired selecting and correlating operations. Such conditioning is often referred to as programming and, with some machines at least, may be accomplished by appropriate wiring of plugboards.

The fourth step—performance of searching and selecting operations—may be simple or complex in nature depending both on the information

¹ It should be noted in this connection that the automatic abstracting of news articles, as published in the weekly magazine *Time*, has been reported recently by H. P. Luhn, “The Automatic Creation of Literature Abstracts,” *IBM Journal of Research and Development*, 2, No. 2, 159–165 (April 1958). For a report of results of application of the same techniques to abstracting scientific review papers as published by the *Sci. American*, see H. P. Luhn, paper presented at the symposium on documentation, School of Library Science, University of Southern California, Los Angeles, April, 1958.

requirement and on the operational characteristics of the equipment being used. Even with relatively simple equipment, such as hand-sorted punched cards, it is instructive to formulate searching operations on a logical basis. (See the next chapter for details.) With fully automatic searching selectors, such formulation is a virtual necessity for efficient programming. For concise expression of such formulations for programming, the Boolean notation is convenient and easy to apply, as discussed subsequently in this chapter and elsewhere.²

From what has been said above, it is perhaps obvious that the ability to perform searching, selecting and correlating operations is dependent on, and determined by, the preliminary analysis of input material, i.e., its encoding. If a high degree of reliability in accomplishing the desired selecting and correlating operations is required, then corresponding care must be devoted to achieving consistency during the analysis-encoding steps. Furthermore, the higher the required degree of selectivity and of correlating ability the more care must be devoted to establishing the rules for analysis of information and for encoding the results of such analysis. Selectivity is determined, not so much by the complexity of the system for analysis and encoding, but rather by the care with which such a system is designed. Each element of complexity in the system must justify itself by providing a useful contribution to advantageous selectivity.

Investigation of the relationships between various forms of complexity, on the one hand, and useful selectivity, on the other hand, is still going forward and is still far from complete. At one time, not so long ago, it was necessary to rely almost entirely on judgment and hunch in establishing systems for analyzing and encoding information for mechanical and electrical selection. This situation has improved considerably during recent years, thanks both to the continuing efforts of a wide circle of workers and also to their willingness to share their experience as written up, for example, in the chapters of this book. By taking such experience into account, it is possible to avoid many pitfalls. It is the purpose of this chapter to summarize such experience and thus provide guide lines for exercising judgment in developing systems for analyzing and encoding information for searching by mechanical and electrical means.

Coding—Its Nature and Role

As already noted, the purpose of preliminary analysis and encoding of information is to make it possible to apply various mechanical and elec-

² See, J. W. Perry, Allen Kent, and M. M. Berry, "Machine Literature Searching," Chapter 6, New York, Interscience Publishers, Inc., 1956; Jessica Melton, and J. W. Perry "Analysis of Questions," Chapter 15 in J. W. Perry and Allen Kent, "Tools for Machine Literature Searching," New York, Interscience Publishers, Inc., 1958.

trical devices to search, to select and to correlate the information contained in a given collection of graphic records. To this end, the subject contents of graphic records are analyzed with respect to those features which can serve as the basis for performing searching, selecting and correlating operations by means of various devices or machines. The results of such analysis must be recorded in a form that enables the searching, selecting and correlating equipment to respond to individual features of the subject contents of records and to combinations of such features. Depending on the equipment used for searching, selecting and correlating this recording assumes a variety of forms. With hand-sorted punched cards, notches are cut at predetermined positions along the cards' periphery. With "Uniterm" cards, document numbers are written or otherwise posted on cards each of which corresponds to some one subject heading (see Chapter 7). With magnetic tape, invisible patterns of magnetic spots are produced. With "Minicards"³ photographic procedures are used to produce patterns of opaque and transparent spots.

These examples of different modes of recording may suffice to make the point that the recording of features of subject matter is in a form whose purpose is to permit mechanical or electrical devices to perform selecting and correlating operations. This form of recording is quite different from printing, writing, diagramming, etc., to which we are accustomed. Recorded material for mechanical or electrical searching is, accordingly, often spoken of as "coded." The term "code" is often applied to rules for generating recordings for automatic or semi-automatic searching. The more complex codes that may be used to advantage with certain types of automatic equipment are sometimes termed "machine language."

Codes may range, therefore, from lists of words or terms for use with simple devices (e.g., hand-sorted punched cards, "Uniterm" cards) to relatively elaborate and carefully designed artificial languages, which resemble machine Esperanto and into which statements in human languages may be interpreted. An essential part of a code is a set of rules for recording selected features of subject matter, e.g., by punching cards, or magnetizing spots on tape, so that one device or another may be used for selecting and correlating operations.

It is perhaps possible to accomplish encoding in such a way that it may appear to be a single step. With a simple device, such as hand-sorted

³ A. W. Tyler, W. L. Myers and J. W. Kuipers, "The Application of the Kodak Minicard Equipment to Problems of Documentation," *Am. Doc.*, 6, 18-30 (1955). cf. also: "Minicard Demonstration," *ibid*, 258-259; "A Minicard System for Documentary Information," Chapter 27 in J. H. Shera, Allen Kent and J. W. Perry (eds.), "Information Systems in Documentation," New York, Interscience Publishers, Inc., 1957.

punched cards, encoding is sometimes practiced by inspecting the information, e.g., reading an abstract typed or mounted on the card, and immediately proceeding with the punching of the card. Even in such a case, certain punching positions must be assigned meaning and this assignment must be kept in mind while encoding. The possibilities of making mistakes and the desirability, not to say necessity, of consistency in encoding make it advisable, even with simple codes, to proceed in a more formal and deliberate fashion, to carry out the following sequences of steps, and consciously to distinguish between them:

1. To decide what features of subject matter shall be expressed by coding.
2. To record such decisions as to important features in a well-defined, orderly fashion. (Preparatory to coding hand-sorted punched cards, for example, an orderly list of important subject headings may be prepared. When applying a comprehensive "machine language," it may be continually extended to include new terms and concepts as they are encountered in papers and reports undergoing encoding⁴.)
3. To decide how each feature will be recorded for mechanical or electrical searching and selecting. (With hand-sorted punched cards, for example, this type of decision will involve assigning some one punching position or some combination of positions to record a given feature of subject matter as expressed, for example, by a corresponding subject heading.)

In this way a set of rules or procedures are established and followed for accomplishing the analysis and encoding of information.

It is, perhaps, instructive to point out certain similarities with subject indexing. Decisions must be made as to what features of subject matter are of sufficient importance to be expressed by coding or by index entries. It is essential both for good indexing and good coding that these decisions shall be made in a consistent fashion in reviewing the subject contents of successive documents. It is equally important that the results of decisions as to important features of the subject contents of documents shall be recorded in an unambiguous and consistent fashion. In subject indexing this leads to the requirement that terminology shall be used in a consistent fashion in setting up index entries. To meet this requirement, lists of carefully compiled, well-defined subject headings are often used when indexing. The same purpose may be served, especially when coding hand-sorted punched cards, by lists of subject headings which also indicate how the cards are to be punched to record the individual items. (See, for example, Chapters 4, 5, 8, 14). With fully automatic searching selectors, simple lists

⁴ See J. W. Perry, Allen Kent and M. M. Berry, "Machine Literature Searching," Chapter 11, New York, Interscience Publishers, Inc., 1956; Jessica Melton and J. W. Perry, Chapter 5, in J. W. Perry and Allen Kent, "Tools for Machine Literature Searching," New York, Interscience Publishers, Inc., 1958.

of terms are not the most effective means for accomplishing the analysis and encoding of recorded information. But with such selectors it remains true, of course, that important features of the subject contents of documents must be recorded in an unambiguous and consistent fashion. To this end, systematic rules, whose aggregate is sometimes termed "machine language," have been developed.⁴

The development of an appropriate code—in other words, a set of rules for analyzing subject matter and recording the results of such analysis—is no simple matter. This statement remains true, even when the code is being developed for such relatively simple devices as hand-sorted punched cards. The cards—rather obviously—are incapable of thinking. Rather they are devices for performing certain operations which can be conducted so as to accomplish useful selections. Similarly, the most advanced searching selectors perform certain operations—albeit more complex in nature—that accomplish useful selections of a more versatile type.

It is perhaps evident that code development must take into account the nature of such selecting operations that can be performed with the aid of various mechanical and electrical devices.

The various devices with which this book is concerned permit, in one fashion or another, a number of characteristics of a given paper, patent, document or graphic record to be independently searched. As a consequence, searching and selecting operations may be directed to any one of the characteristics so recorded. This capability, of itself, is not sufficiently useful or interesting to insure widespread application of searching equipment no matter how flexible or rapid in operation.⁵ The effectiveness of all kinds of searching and selecting equipment—from hand-sorted punched cards to fully automatic electronic selectors—results, rather, from their ability to perform searching operations that are defined in terms of combinations of characteristics. Some simple considerations suffice to indicate the underlying principles.

Selecting Operations and Equipment Capabilities

As pointed out in various chapters in this book and elsewhere,⁶ a considerable variety of mechanical and electronic devices have been applied

⁴ Ralph R. Shaw, "Machines and the Bibliographical Problems of the Twentieth Century," in Louis N. Ridenour, Ralph R. Shaw and Albert G. Hill, "Bibliography in an Age of Science," Urbana, University of Illinois Press, 1952; Carl S. Wise and J. W. Perry, "Multiple Coding and the Rapid Selector," *Am. Doc.*, 1, 76-83 (1950).

⁵ Marjorie R. Hyslop, "Inventory of Methods and Devices for Analysis, Storage and Retrieval of Information," Chapter 6 in J. H. Shera, Allen Kent and J. W. Perry (eds.), "Documentation in Action," New York, Reinhold Publishing Corporation, 1956.

to achieve advantageous selection and correlation of recorded information. To characterize their capabilities and limitations in an exhaustive fashion would require a lengthy treatise. In spite of obvious differences in mode of functioning, such diverse devices as hand-sorted punched cards and electronic selectors have a number of important operational characteristics in common.

1. Independent recording of characteristics of the subject contents of documents. (The range and nature of characteristics that can be conveniently and advantageously recorded does vary, however, depending on the device or equipment being used.)

2. Searching and selecting operations may be directed to any one of the independently recorded characteristics or to their combinations. (The range of such combinations depends not only on the extent to which different types of characteristics may be recorded but also on the ability of various searching and selecting devices and equipment to detect various characteristics and to respond to them.)

3. Concise statement of searching and selecting operations may be accomplished with the notation of Boolean algebra. (Chapter 19 indicates how the Boolean notation may be applied to specifying various selecting operations performed with hand-sorted punched cards. Similar formulations are also applicable to aspect cards, e.g., "Peek-a-boo," "Uni-term," and the like. Furthermore, the same mode of abstractly specifying selecting operations has also been used both in designing fully automatic equipment and in formulating procedures for encoding abstracts.⁷)

Subsequent discussion may serve not only to show how the performance characteristics of various kinds of equipment may be formulated, but also to indicate the wide range in capabilities and limitations of present-day devices. Selection of a given device or combination of devices to meet a given situation can be expected to require its careful analysis in terms of purposes to be served, the value of benefits that can be provided and the costs of conducting various operations, both intellectual and routine in nature.⁸

With punched cards, for example, individual holes may be punched to

⁷ For a summary of underlying principles, see: J. W. Perry, Allen Kent and M. M. Berry, "Machine Literature Searching," Chapters 11 and 13, New York, Interscience Publishers, Inc., 1956; J. W. Perry and Allen Kent, "The New Look in Library Science," *Appl. Mechanics Revs.*, 9, No. 11, 457-60 (1956). See also, for fully detailed presentation: J. W. Perry and Allen Kent, "Tools for Machine Literature Searching," New York, Interscience Publishers, Inc., 1958.

⁸ For an introductory discussion of a general theory for analysis of costs and benefits, see J. W. Perry and Allen Kent, "Documentation and Information Retrieval," New York, Interscience Publishers, Inc., 1957, cf., also Chapter 9 in J. W. Perry, Allen Kent and M. M. Berry, "Machine Literature Searching," New York, Interscience Publishers, Inc., 1956.

designate various subject matter characteristics. Alternately various combinations of holes may be punched so that some combination denotes a certain characteristic as designated when establishing the code. A search for needed information may then be formulated in terms of various characteristic combinations each of which is recorded either by punching a single hole (direct coding) or by punching a combination of holes (more complex codes as exemplified by numerical codes, random superimposed codes, and the like). In discussing such selecting operations, and in formulating them concisely, it is convenient to use capital letters (A, B, C, D, etc.) to designate, with hand-sorted punched cards, a certain hole or combination of holes to which meaning has been assigned. More generally, capital letters, with subscripts, may be used to designate meaningful recorded patterns and their combinations at various levels corresponding, for example, to "words," "phrases," "sentences," and the like, in machine language for use with fully automatic selectors as summarized previously in chapter 11.

One of the simplest, and most important forms of combinations of characteristics may be expressed as the search requirement that all of several (two or more) characteristics shall be detected. With hand-sorted punched cards, for example, a search may be conducted by selecting out those cards that have all of several (two or more) positions (or meaningful combinations of positions) punched so as to respond to sorting operations. As discussed in detail in Chapter 19, such sorting operations may be designated symbolically as

$$A \cdot B \cdot C \cdot D - - -$$

where the individual letters indicate the various positions that may be punched in each of the cards, so that those cards that are punched may be separated from others that are not so punched. With aspect cards, a similar search may be conducted by observing those document numbers that are present on all of several (two or more) of the aspect cards.

Similarly, we may select out those cards that have been punched at any one, at least, of several punching positions. Then we may select those cards that have been punched at position A and/or position B, and/or position C, and/or position D, etc. Such a sorting operation is conveniently and concisely symbolized by the logical sum, an example of which follows:

$$A + B + C + D, \text{ etc.}$$

A third possibility is to reject those cards that are punched in a given position. For example, we may reject those cards punched in position B while accepting those punched in position A. Such a specification of search is an example of a logical difference and may be symbolized by $A - B$.

It is theoretically possible to carry out, with hand-sorted punched cards—see Chapter 19—(or aspect cards), more complex sorting (or matching) operations of any degree of complexity which involve two or more of the three above mentioned logically defined basic operations. Such complex sorting (or matching) operations may be exemplified as follows:

$$\begin{aligned}(A + B) (C + D) \\ (A \cdot B) + (C \cdot D) \\ [(A \cdot B) + (C \cdot D)] [(E - F)] \\ (E - F) + (C \cdot D)\end{aligned}$$

As performed with hand-sorted punched cards, the logically defined sorting operations relate to the holes and notches cut in the cards. These operations, strictly speaking do no more than separate the cards that are characterized by certain holes and notches. With aspect cards, the logically defined number-matching operations do no more than identify certain documents that are characterized by having their serial numbers or similar identification recorded on the aspect cards involved in a particular search.

The practical limitations with regard to complexity of logically defined operations that can be performed conveniently show up in different forms and degrees, depending upon the type of searching device or principle used. With hand-sorted punched cards, the number of cards to be “needled” as the size of the file increases usually provides the first trials of patience in conducting manual operations. With aspect cards, the first trials of patience show up as the complexity of searching operations increase, particularly with regard to the conducting of selecting and correlating procedures that involve logical products one or more of whose terms is a logical sum.

Assigning meaning to the holes and notches of punched cards (or to the cards of the aspect system) is an entirely independent operation which is, of course, essential to purposeful use of the cards to accomplishing useful searches, selections and correlations.

When meaning is attributed to each individual punched position, the form of coding is termed “direct coding.” Meaning may also be ascribed to each of various combinations of holes in a wide variety of ways. The two main types of such combinational codes employed with hand-sorted punched cards may be distinguished as follows:

1. In a given field, (i.e., in a given group or set of holes) only one combination may be punched in any one card. (But, of course, within the same field, different combinations may be punched in different cards. Examples of this form of coding are numerical and alphabetical codes described in Chapter 2, pages 18–21.)

2. In a given field, in any one card, more than one combination may be punched. (The number of such combinations that may be advan-

tageously punched in any one card will, however, be limited by the possibility of undesired fortuitous generation of false or "ghost" combinations and by the accompanying possibility of cards being selected on the basis of "ghost" combinations that do not correspond to intended code entries for such cards⁹. To counteract the tendency to generate false or "ghost" combinations in this form of coding, random numbers are often used to establish each of the combinations to which meaning is assigned. From these practices, the name "random superimposed coding" has been derived.)

Standard coding for machine manipulated cards calls for punching one or more holes in a column of a card to designate a single letter, numeral or other symbol. (See Chapter 3, pages 55, 65). Both the direct and random superimposed form of coding have also been applied to machine manipulated cards which may be sorted and selected by various machines including relatively low-cost sorters, as available, for example, from IBM or Remington-Rand.

Coding for aspect systems using machine manipulated cards may involve punching document numbers, one to a card, on separate decks of cards, each deck representing a code, a word, or an idea. Identification of desired documents may be effected either by automatic routines for collating the decks of cards representing each of the words or ideas involved in a search, or by visually identifying number matches.

The mechanical sorters used in both of these approaches are so designed and constructed that advantageous performance, in a practical sense, of a searching operation directed to a given hole (or combination of holes) requires that it be known in which column the hole (or combination of holes) in question has been punched. This remains true when a sequence of columns (or "field") is used to record no more than a single meaningful combination of holes in any one card or when, by applying random superimposed coding, a plurality of meaningful combinations are punched in a given "field" of a single card. With both these approaches to coding, the columns are organized into rigidly defined sets or "fixed fields" as they are usually termed.

Coding based on fixed fields is subject to severe limitations, as was understood nearly a decade ago¹⁰. In particular, with a limited number of fixed fields, it is necessary, for reliable operation, to establish a corresponding organization of the subject characteristics of the information to

⁹ See Carl S. Wise, "Mathematical Analysis of Coding Systems," Chapter 20 (in particular pp. 285-299) in Robert S. Casey and J. W. Perry (eds.) "Punched Cards. Their Application in Science and Industry," New York, Reinhold Publishing Corporation, 1951. Revised and reprinted as Chapter 21 of this book.

¹⁰ J. W. Perry, "The ACS Punched Card Committee. An Interim Report," *Chem. Eng. News*, 27, 754-756 (1949). cf., also *ibid*, 28, 3789 (1950).

be encoded. Similarly, for reliable as well as efficient operation of an aspect system, the requirement to devote a single physical record to an aspect representing an area of subject matter makes it necessary to make arbitrary decisions for controlling the scope or range of subject matter covered by each aspect record. With certain kinds of information it is relatively easy to organize the important characteristics into mutually exclusive sets with one characteristic in each of the sets being both necessary and sufficient for a given item of information. If, for example, the item of information is a business transaction such as the selling of a certain shipment of some one product, then the various sets of characteristics will be typified by (i) the name, code number or other designation of the product (ii) the amount sold (iii) the unit price (iv) the customer to whom sold (v) the salesman involved (vi) date of sale (vii) date of delivery, and so on. One characteristic each from such sets provides the needed information regarding the sale, and assignment of appropriate fields on the punched cards enables the various characteristics to be recorded in the preassigned fixed field. Such sets of characteristics are spoken of as "mutually exclusive" for the reason that specification of any one characteristic within a given set excludes the possibility of another characteristic within the same set being pertinent to a given item of information. Thus, specification of one unit price for a given shipment of product excludes the possibility that another unit price should apply to the same shipment. The sale to one customer excludes simultaneous sale of the same shipment to another customer. The delivery on one date excludes delivery of the same shipment on another date, etc. It should also be noted that these sets of characteristics are also mutually independent. Thus the unit price of a product may change independently of the amount sold, or the customer may vary independently of the date of sale, etc.

In many practical cases of characterizing recorded information some of the characteristics may be organized into mutually exclusive sets, while other characteristics are not amenable to such treatment. For example, personnel data relating to various individuals involves certain sets of characteristics that are mutually exclusive. Thus, birth at one date excludes the possibility of birth on some other date. Similarly, birth at some one town or other location excludes birth somewhere else. Thus, birth date and place of birth constitute examples of sets of characteristics that are mutually exclusive. On the other hand, knowledge of one language does not exclude the possibility of knowledge of one or more other languages. Similarly, attendance at one university or college does not exclude the possibility of attendance at another being included in personal data. Attempting to fit such characteristics as knowledge of languages and attendance at universities or colleges into a fixed field scheme is sure to lead

to difficulties. When the subject contents of documents pertain to a wide range of characteristics that cannot be organized into a restricted number of mutually exclusive sets, the fixed field system of machine-sorted punched cards presents difficulties which become more severe as the range of such characteristics becomes wider. Even with moderately broad ranges of characteristics the attendant difficulties impose limitations whose surmounting justifies considerable effort.

As pointed out in Chapter 2 (pages 55, 65) fixed field coding with machine-sorted punched cards is usually carried out by recording a single symbol, especially some one numeral or letter, in a single column of the card. With IBM cards, for example, this means that the twelve punching positions in a column are being used to record a single symbol. Direct coding—i.e., assignment of a definite well-formulated meaning to each individual hole in a specified field in a machine-sorted punched card—is a simple form of coding which enables more than one entry to be made in one column. This method has been applied with impressive success in dealing with narrow ranges of subject matter¹¹. But it is now widely recognized that this approach is limited in the range of subject matter that may be encoded in sufficient detail to provide discriminating and correlating capabilities such as are needed in dealing with extensive files of complex information as encountered, for example, in science and technology. Another way in which the limitations of fixed fields may be partially surmounted is by the use of random superimposed coding. (See Chapter 10.) With this method, a restricted number of characteristics within a given set may be recorded in a given card in the fixed field reserved for the set of characteristics. The restriction in number of such characteristics that may be recorded on a single card will depend on the number of punching positions in the field, the number of holes in a meaningful combination, the extent to which false sorts may be tolerated and related factors.

In aspect systems, the intellectual analog of the fixed field coding problem is encountered because of the operational requirement to organize characteristics into sets, in order to achieve practical results during the exploitation phase. Because of the practical limitations encountered in aspect systems in performing searching and correlating operations involving logical products of series of logical sums, the need to reduce scatter of synonymous or partially synonymous terminology which may be used as headings of

¹¹ Julius Frome and Jacob Leibowitz, "A Punched-Card System for Searching Steroid Compounds." Patent Office Research and Development Reports. No. 7, Washington, Dept. of Commerce, July 8, 1957. Don. D. Andrews. "Progress Report on U. S. Patent Office Mechanized Searching" Paper presented before the Division of Chemical Literature, American Chemical Society, 113th National Meeting, San Francisco, California, April, 1958.

aspect cards becomes controlling. This consideration underlies the coding described in Chapters 6 (pages 141-146) and 7.

In order to go further in surmounting the limitations of fixed field coding so that broader ranges of subject matter could be rendered amenable to mechanical or electrical searching and selecting operations, various machines have been specially designed in recent years. The Luhn Scanner was designed to pass IBM cards end-wise and to detect various letters or other symbols as distinctive combinations of five holes in the twelve punched positions within any one column. The machine detected individual symbols or meaningful sequences of symbols so recorded in a fashion analogous to a blind man reading Braille and detecting an individual letter, e.g., "c", or a letter sequence, e.g., "c—a—t." The ability of the machine to detect combinations of meaningful symbols or their sequences as exemplified by words or codes was restricted, however, to a single level of logical combinations.¹²

Recent machine development work has surmounted this limitation. It has been found possible to construct—at a cost not exceeding a fraction of the price of a general-purpose computer—searching equipment¹³ characterized by the following capabilities:

1. Symbols and symbol sequences may be entered in the recording medium, e.g., magnetic tape, without limitations of the fixed field type. Symbols and symbol sequences are detected as code patterns of the same type as used with Teletype tape or the above mentioned Luhn Scanner.

2. Symbols and symbol sequences may be organized into combinational levels analogous to "syllables," "words," "phrases," "sentences," "paragraphs," etc. Letters with subscripts provide a simple notation for distinguishing such levels, thus we have:

A_1, B_1, C_1, D_1 , etc., for "syllables"

A_2, B_2, C_2, D_2 , etc., for "words"

¹² Staff report, "Machine Techniques for Information Selection," *Chem. Eng. News*, 30, 2806-10 (1952); H. P. Luhn, "The IBM Electronic Information Searching System," International Business Machines Corp., Poughkeepsie, New York, 1952.

¹³ J. W. Perry, "The Western Reserve University Searching Selector," Chapter 18 in J. W. Perry and Allen Kent "Tools for Machine Literature Searching," New York, Interscience Publishers, Inc., 1958. Note also that recent studies and tests have demonstrated that appropriate programming enables at least certain electronic computers, in particular the IBM 650 and the IBM 705, to accomplish searching and selecting operations as specified above. See Sally F. Dennis, "Programming the IBM 650 to Searching Encoded Abstracts," Chapter 19 in J. W. Perry and Allen Kent "Tools for Machine Literature Searching," New York, Interscience Publishers, Inc., 1958.

A_3, B_3, C_3, D_3 , etc., for "phrases"

A_4, B_4, C_4, D_4 , etc., for "sentences"

A_5, B_5, C_5, D_5 , etc., for "paragraphs"

A_6, B_6, C_6, D_6 , etc., for "messages"

3. At any level, which we may term the "n-th" level, an encoded combination will consist, as a rule, of a number of component combinations at the " $n - 1$ " level. Each of several "n-th" level combinations, denoted by A_n, B_n, C_n, D_n , etc. may be specified in terms of component elements designated by $A_{n-1}, B_{n-1}, D_{n-1}$, etc., or more generally in terms of lower level component elements exemplified by, $A_e, B_j, A_k, C_m, D_e, R_f, A_f, C_e$, etc., where j, k, m, e and f denote levels lower than n . As discussed in detail in Chapter 11 pages 252-256, combinations of elements may be specified at any level in terms of logical product, sum and difference and their complex combinations.

The construction of moderate-cost equipment having the above outlined identifying and selecting capabilities has made it possible to express the essential contents of scientific and technical papers in the form of encoded abstracts, whose syntax is a simplified standardized system for recording relationships and whose component terms are encoded to express generic aspects of meaning¹⁴. Such a system can be applied to a broad field such as metallurgy. The more general application of this type of methodology opens up the possibility of establishing a network of agencies that would analyze, abstract and encode information in different fields on the basis of methodology based on common logical principles. Such a network—resembling somewhat the present world-wide telephone network—would greatly facilitate the interdisciplinary exchange of information and it would thus open up new perspectives for efficient use of scientific and technical information¹⁵.

Principles of Code Development

As already noted, the operations performed by, or with the aid of, various mechanical and electrical devices can be formulated by abstract expressions that are mathematical in nature. These operations, when so formulated, are, of themselves, as devoid of meaning as any other system of symbols or signs. They achieve meaning as a consequence of deliberate assignment of significance. This step is, rather obviously, of key impor-

¹⁴ See footnote 7, page 399.

¹⁵ For a more detailed discussion, see: J. W. Perry, Allen Kent and M. M. Berry, "Machine Literature Searching," Chapter 8, New York, Interscience Publishers, Inc., 1956; also, J. W. Perry and Allen Kent "Documentation and Information Retrieval," Chapter 3, New York, Interscience Publishers, Inc., 1957.

tance in determining the degree of benefit in applying mechanical and electrical devices to the searching, selection and correlation of recorded information.

Appropriate assignment of meaning to certain forms of recording, e.g., the punches in cards, or the magnetic spots on computer tapes, does not, of course, accomplish the analysis of the subject contents of documents. Such analysis is equally necessary as a prerequisite to successful application of mechanical or electrical devices. Consistency in such analysis is, rather obviously, not assured by the fact that the selecting operations performed with the aid of various devices may be precisely formulated in terms of the Boolean notation.

Appropriate assignment of meaning to certain forms of recording, i.e., the development of a specific code, must take into account, on the one hand, the nature of the selecting operations that can be performed by various types of devices or equipment and, on the other hand, the scope and the nature of the subject matter to be analyzed and encoded.

Let us next direct attention to certain important consequences of the nature of the selecting operations that can be performed with various devices.

Many different coding methods and different systems for applying various mechanical and electrical devices, as described in this book and elsewhere, make extensive use of selection based on the logical product. There is a fundamental reason for this which may be demonstrated as follows.

Let us assume, for simplicity, that the subject contents of a file of documents can be analyzed in terms of 100 characteristics of which, on an average, ten will apply with equal probability to any one individual document. Then, a search for all documents pertaining to one characteristic, e.g., recorded as A, will result in one-tenth of the documents being selected. If, now, a search is directed to all documents pertaining to both of two subjects, e.g., to the logical product of the recordings $A \cdot B$, then one-hundredth of the documents will be selected. For a three-term logical product, e.g., $A \cdot B \cdot C$, one-thousandth, i.e., 0.1 per cent of the file will be selected. If we remember that we have assumed that the analysis of the subject contents of documents results in each code characteristic, recorded as A, B, C, D, etc., corresponding precisely to what the searcher would like to find under that heading, we may summarize this ideally favorable situation as shown in Table 18-1. Evidently searches defined in terms of logical products can permit us to achieve elimination of rapidly increasing proportions of the original file as the number of terms in the logical product is increased. In achieving this purpose with maximum efficiency, it is important that the terms in the logical product shall partake of the nature of independent variables. When this condition is not fulfilled, as in an

TABLE 18-1. SELECTIVITY INCREASE WITH INCREASE IN TERMS IN LOGICAL PRODUCT
(Assuming Perfect Correspondence between Analysis and Search Requirement)

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Parti- ent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	10	100	0	0
2	1	100	0	0
3	0.1	100	0	0
4	0.001	100	0	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	100(0.1) ⁿ	100	0	0

extreme case when all items coded for A are also coded for B, it is clear that a search directed to A·B will be no more selective than a search directed to either A or B above.

The preceding paragraph presents one reason why considerable careful thought must be devoted to selecting those characteristics that are to serve as the terms in logical products for searching and selecting purposes. In the discussion of logical products that follows immediately and pertains to Tables 18-1 through 18-6, it will be assumed that the characteristics are not interdependent. In other words, the fact that some one characteristic pertains to the subject contents of a given document is without influence on the probability that any other of the characteristics in our hypothetical code pertain to the subject contents of the same document.

In selecting characteristics to construct a code, another feature of logical products must also be taken into account. This relates to the possibility that a search directed to one or more characteristics does not result in selection of precisely those documents that are of pertinent interest to a given information requirement. There are three important possibilities:

(1) Such a search does not result in selection of all pertinent information, but no non-pertinent information is selected.

(2) Such a search selects all pertinent information but it also selects additional information not of pertinent interest.

(3) Such a search does not result in selection of all pertinent information but it also results in selection of additional information not of pertinent interest.

In the first case, by way of example, let us return to our previous example and assume that for any one characteristic, such as those recorded by A, B, C, etc., 9 per cent of the documents in the file is selected rather than 10 per cent as before. In this case, we will assume that the 9 per cent selected is all pertinent and that the remaining 1 per cent corresponds to pertinent information not selected but overlooked and consequently

TABLE 18-2a. INCREASE IN LOSS OF INFORMATION WITH INCREASE
IN TERMS IN THE LOGICAL PRODUCT
(Assuming each characteristic retrieves 90% of the information it should)

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	9	90	10	0
2	0.81	81	19	0
3	0.0729	72.9	27.1	0
4	0.00656	65.6	34.4	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.09)^n$	$100 \left(\frac{0.09}{0.1} \right)^n$	$100 - 100 \left(\frac{0.09}{0.1} \right)^n$	0

TABLE 18-2b. INCREASE IN LOSS OF INFORMATION WITH INCREASE
IN TERMS IN THE LOGICAL PRODUCT
(Assuming each characteristic retrieves 99% of the information it should)

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	9.9	99	1	0
2	0.98	98	2	0
3	0.097	97	3	0
4	0.0096	96	4	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.99)^n$	$100 \left(\frac{0.099}{0.1} \right)^n$	$100 - 100 \left(\frac{0.099}{0.1} \right)^n$	0

effectively lost. Under these conditions, the results of searches involving logical products with increasing numbers of terms may be summarized as shown in Table 18-2a. If, for each characteristic in the logical product, 9.9 per cent of the file were selected (instead of 9 per cent) and the information lost by non-selection for any one characteristic were 0.1 per cent (instead of 1 per cent), the results would be more favorable as shown in Table 18-2b.

As mentioned above, it is also possible for a search formulated by a logical product to select all pertinent information and, in addition, information that is not of pertinent interest. Let us assume that each term in the logical product selects 11 per cent of the file, with 10 per cent of pertinent

TABLE 18-3a. INCREASE IN EXTRANEOUS INFORMATION WITH INCREASE IN TERMS IN THE LOGICAL PRODUCT

(Assuming each characteristic retrieves 11% of the file—10% pertinent, 1% extraneous)

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	11	100	0	9.1
2	1.21	100	0	17.4
3	0.133	100	0	24.9
4	0.0146	100	0	31.7
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.11)^n$	100	0	$100 \left[\frac{(0.11)^n - (0.1)^n}{(0.11)^n} \right]$

TABLE 18-3b. INCREASE IN EXTRANEOUS INFORMATION WITH INCREASE IN TERMS IN LOGICAL PRODUCT

(Assuming each characteristic retrieves 10.1% of the file—10% pertinent, 0.1% extraneous)

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	10.1	100	0	0.99
2	1.02	100	0	1.96
3	0.103	100	0	3.00
4	0.0104	100	0	3.86
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.101)^n$	100	0	$100 \left[\frac{(0.101)^n - (0.1)^n}{(0.101)^n} \right]$

interest and 1 per cent extraneous and non-pertinent. In this case, results as shown in Table 18-3a would be obtained. If, for each characteristic in the logical product, 10.1 per cent of the file were selected (with 10 per cent pertinent information and 0.1 per cent not pertinent), then the results, as shown in Table 18-3b would be more favorable.

It is, of course, quite possible that a selecting operation directed to a given characteristic may result in selection of extraneous information while, simultaneously, all pertinent information is not retrieved. For

TABLE 18-4a. INCREASE BOTH IN LOST INFORMATION AND IN EXTRANEOUS INFORMATION WITH INCREASE IN TERMS IN LOGICAL PRODUCT

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	10	90	10	10
2	1	81	19	19
3	0.1	72.9	27.1	27.1
4	0.01	65.6	34.4	34.4
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.1)^n$	$100 \left(\frac{0.09}{0.1} \right)^n$	$100 - 100 \left(\frac{0.09}{0.1} \right)^n$	$100 \left[\frac{(0.1)^n - (0.09)^n}{(0.1)^n} \right]$

TABLE 18-4b. INCREASE BOTH IN LOST INFORMATION AND IN EXTRANEOUS INFORMATION WITH INCREASE IN TERMS IN LOGICAL PRODUCT

Number of Terms in Logical Product	Per cent of Total File Selected	Per cent of Pertinent Information Selected	Per cent of Pertinent Information Lost by Non-selection	Per cent of Information Selected that is not Pertinent
1	10	99	1	1
2	1	98	2	2
3	0.1	97	3	3
4	0.01	96	4	4
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	$100(0.1)^n$	$100(0.099)^n$	$100 - 100 \left(\frac{0.099}{0.1} \right)^n$	$100 \left[\frac{(0.1)^n - (0.099)^n}{(0.1)^n} \right]$

example, let us assume that a search of any one characteristic results in selecting 10 per cent of the file of which 10 per cent of the pertinent information constitutes 9/10 and the non-pertinent information 1/10. In this case, increasing the number of terms in the logical product leads to the results shown in Table 18-4a. If, in the case under consideration, the proportion of pertinent information selected by any one characteristic increases from 9/10 to 99/100 and, correspondingly, the proportion of extraneous information decreases from 1/10 to 1/100, then much more favorable results are obtained as shown in Table 18-4b.

In working out numerical examples to illustrate the various effects of increase in number of terms in a logical product on search results, the calculations have been kept simple by assuming that for each of the special cases investigated, all the characteristics used to analyze the subject

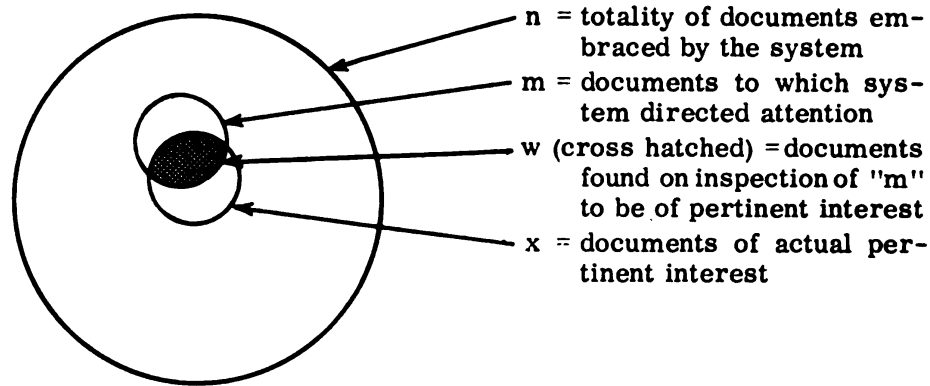


Figure 18-1. Diagram exemplifying general case of relationship between n , m , x and w .

contents of documents are functionally effective in the same way and to the same degree. Thus in the calculations for Table 18-4a, it was assumed that each characteristic resulted in selection of 10 per cent of the file and that 9/10 of the selected items would be of pertinent interest. In actual situations, such uniformity of functional effectiveness of all characteristics can scarcely be anticipated. If we assume differences, however, the character of the results remains unchanged. This is true, in particular, of the trends observed as the number of terms in the logical product are increased. To illustrate this point, the diagram shown in Figure 18-1 is helpful. Here the area within the circle n indicates the totality of documents in a given file while the area of the circle m indicates those documents which are selected by a search directed to a given characteristic, as recorded by code A. The area of circle x indicates those documents that a person making use of the system would like to have brought to his attention by a search directed to the encoded characteristic A. The shaded area, w , indicates those documents of pertinent interest that are included among the documents selected as encoded for a given characteristic. The circles m and x will be congruent, if, and only if, there is complete agreement between the person who does the encoding and the person who needs information as to which documents should be characterized by the encoded characteristic A. In general, the degree of agreement will deviate, more or less, from 100 per cent and various cases of deviations from complete agreement have been discussed elsewhere¹⁵.

In previous discussion, it was emphasized that the areas within the various circles in diagrams such as Figure 18-1 and also the degree of overlap of circles m and x are independent variables, which are of decisive importance in determining the operational effectiveness of an information system when it is used to select documents from a file or library. To illustrate this important point further, let us assume that we have an

TABLE 18-5a. DATA TO ILLUSTRATE SEARCH RESULTS
WITH DIFFERENT CHARACTERISTICS

<i>Characteristic</i>	<i>n</i>	<i>m</i>	<i>x</i>	<i>w</i>
A	15000	300	250	240
B	15000	1500	1515	1490
C	15000	750	1000	700
D	15000	160	156	140

TABLE 18-5b. OPERATIONAL SEARCH FACTORS WITH DIFFERENT CHARACTERISTICS

<i>Characteristic</i>	<i>m/n</i>	<i>x/n</i>	<i>w/x</i>	<i>w/m</i>
A	2%	1.667%	96.0%	80%
B	10%	10.1%	98.35%	99.4%
C	5%	6.67%	70%	93.4%
D	1.07%	1.04%	89.75%	87.5%

information system in which four characteristics perform, in a qualitative fashion for various interrelated searches, as shown in Figure 18-1 and that the quantitative performance provides data as given in Table 18-5a. The corresponding values for m/n , x/n , w/m and w/x are given in Table 18-5b for the four subject matter characteristics A, B, C, D. When logical products are set up among these four characteristics to define the scope of searches, changes must be anticipated in the per cent of pertinent information retrieved (and also the per cent of pertinent information lost by non-selection) as well as in the per cent of selected information that is pertinent (and also the per cent of selected information that is not pertinent). These changes, as shown in Table 18-6 were calculated in the same way as the data presented in Tables 18-1 through 18-4. The data in Table 18-6 display the same trends as to changes in reliability of selection of pertinent material with increasing number of terms in the various logical products, as were evident from the previously presented tables.

The preceding discussion of the functional effects of logical products in searching operations leads to the following two conclusions as to how characteristics of subject matter are to be selected in constructing codes, if searches defined in terms of logical products are to achieve optimum selectivity and reliability.

1. Selectivity in searching, that is, reduction in the per cent of the file selected with increasing number of terms in a logical product, is at a maximum when such terms correspond to characteristics that are independent variables in the sense that the relevancy of one characteristic to the subject contents of a document does not influence the probability that one or more of the other terms is also relevant.

2. The degree of reliability of searching operations, that is the probability of selection of documents of pertinent interest (and the probability of rejection of those that are not pertinent), depends on the extent to which

TABLE 18-6. CHANGES IN RELIABILITY OF SELECTION OF PERTINENT MATERIAL WITH INCREASE IN NUMBER OF TERMS IN LOGICAL PRODUCT

Per cent of Total File Selected	Per cent of Total File Actually of Pertinent Interest	Per cent of Pertinent Information Retrieved	Per cent of Pertinent Information Lost by Non-selection	Per cent of Selected Information that is Pertinent	Per cent of Selected Information that is not Pertinent
A 2	1.67	96.0	4.0	80.0	20.0
B 10	10.1	98.4	1.6	99.4	0.6
C 5	6.67	70.0	30	93.4	6.4
D 1.07	1.04	89.8	10.25	87.5	12.5
A·B 0.2	0.168	94.5	5.5	79.5	20.5
A·C 0.1	0.111	67.2	12.8	74.7	25.3
A·D 0.024	0.0173	86.2	13.8	70.0	30.0
B·C 0.5	0.673	68.9	31.1	92.8	7.2
B·D 0.107	0.105	88.4	11.6	87.0	13.0
C·D 0.0535	0.0693	62.9	37.1	81.7	8.3
A·B·C 0.01	0.0112	66.1	33.9	74.2	25.8
A·B·D 0.00214	0.00175	84.8	16.2	69.6	30.4
A·C·D 0.00107	0.00113	60.4	39.6	65.4	34.6
B·C·D 0.00535	0.0070	61.9	38.1	81.2	8.8
A·B·C·D 0.000107	0.00117	59.4	40.6	65.0	35.0

the relevancy of various characteristics, as discerned during the analysis of the subject contents of documents, corresponds in scope to the relevancy of the same characteristics to an information requirement.

Let us first consider some of the practical implications of the first of these two conclusions.

In striving for a high degree of selectivity in searches defined by logical products, it is not intended to imply that only those characteristics that are completely independent can be used in conducting such searches. In most practical cases, it would be difficult to achieve complete perfection in this regard. Other considerations being equal, especially clear and unambiguous definability as discussed below, those characteristics that are less interdependent are to be preferred in code construction. This is particularly true when working with those devices, e.g., hand-sorted punched cards, whose design imposes practical limits on the number of characteristics that can be punched in any one card.

In organizing a code, especially for hand-sorted punched cards or for fixed fields on machine-sorted cards, it is often possible to organize characteristics into sets which are, in large measure at least, independent in character. Thus, one such set might constitute a materials index and another set of characteristics might constitute an index of processes and properties. Chapter 5 presents an example of a hand-sorted punched-card code organized along these lines so as to enhance its convenience of use. If

the characteristics within a given set are mutually exclusive in nature, that is to say, if the relevancy of one of the set's characteristic to a given document excludes the possibility that a companion characteristic may be relevant, then a numerical form of code (see Figures 2-9, 2-11, 2-12, pages 20, 21) or an alphabetical code (see Figure 2-10, page 20), by permitting the coding of the set of characteristics in a smaller number of punching positions, may be preferable to other forms of coding, especially direct coding. On the other hand, if the characteristics in a set are rather numerous and if, in addition, they are of such a nature that a small number, in varying combinations, can be expected to be relevant to the subject contents of different documents, then it may be advisable to record them on hand-sorted punched cards by applying the principles of random superimposed coding. These same observations as to probable advantages of numerical, alphabetical and random superimposed coding apply also to machine-sorted punched cards when their selection is carried out on the basis of fixed fields.

Another factor of decisive importance both in constructing codes and in selecting devices and equipment is the number of documents or similar items that are involved or can be expected to become involved in a particular situation. Of decisive importance in this connection is the relationship between number of items on the one hand and practical selectivity requirements on the other hand. Consideration of illustrative numerical data is perhaps the quickest way to make this point. If we are dealing with a file of 10,000 items, and if the selectivity of a search is such that 0.1 per cent of the file is selected, then the search results in 10 items being picked out for personal review. The same degree of selectivity applied to a file of 100,000 items would, of course, result in 100 items for personal review and with a file of 1,000,000 items, the number selected for personal review would increase to 1,000. This latter number of selected documents constitutes a situation that is almost sure to result in attention being directed to the possibilities and advantages of reducing the number of documents to be reviewed personally. In other words, the need will be felt for more sharply selective, better focused searches which in turn will require increased selectivity on the part of the searching system. A degree of selectivity quite acceptable for files of moderate size may prove quite inadequate in dealing with large files. This fact often makes it necessary to exert unusual care and considerable reserve in evaluating the results of small-scale tests and demonstrations which may tend to cause the mass effects of large files to be underestimated or overlooked completely.

These considerations as to the total number of items selected during searching operations also apply to the percentages of extraneous items that may be tolerated when dealing with files of various sizes. If, in the

example just given, it were found that 6 of the 10 items selected from the file of 10,000 were of marginal interest or even completely non-pertinent, the time lost in personally reviewing these six items would not be excessive. It would probably not be advantageous to develop a more effectively selective coding system so that most or all of the 6 extraneous items could be eliminated by mechanically performed selecting operations. With a file of 100,000 items, 100 items would be selected by machine and 60 of them would be found, on personal inspection, not to be pertinent.

Other things being equal if two coding systems are designed with equal skill and care, the more selective is virtually certain to involve greater costs or greater time and effort in the analysis of information and in conducting searches. Justification for incurring increased costs may exist, however, when dealing with larger files. With the 100,000 item file, the amount of time and effort required personally to review 60 non-pertinent documents cannot be regarded as negligible while personal review of 600 nonpertinent items from a 1,000,000 item file is virtually certain to be considered prohibitively time consuming.

As requirements for selectivity become more severe, it is necessary, as has already been emphasized, to take into account the degree of reliability of searching operations, that is to say, the probability of selection of documents of pertinent interest and the probability of rejection of those that are not pertinent. In particular, it is necessary, in order to meet stricter selectivity requirements, to reduce margins of difference between scope of relevancy of a given characteristic during input processing of information and output searching operations. In striving toward this goal, fundamental importance is attached both to the definition of concepts and corresponding terminology as well as to consistency in the use of terminology. Such consistency, for the more specific fields of specialization, may be achieved in adequate measure by conscientious application of carefully formulated codes, which in certain important respects, at least, have much in common with the subject authority lists widely used in conventional library operations. When dealing with broader and broader ranges of subject matter, consistent application of standard terminology becomes increasingly difficult. The burden that would otherwise have to be imposed on persons charged with the responsibility of analyzing the subject contents of documents can be eased in large measure by setting up encoding procedures which, in effect, accomplish the interpretation of terminology on a consistent standardized basis. As a consequence of such encoding, terms that are similar in meaning are interpreted into codes that have corresponding meaningful elements in common.

Selectivity of codes designed for use with large files may also be enhanced in a very advantageous fashion by recording the relationships between such substantive entries as various materials or organisms,

equipment or devices, properties, functions, processes, conditions, and the like. Thus, the role of a given material as "starting material," "final product," etc., may be recorded as a characteristic of the subject contents of a given document. Similarly, the relationships of a given entity and its properties, and a process and its attendant conditions, may serve as a means for enhancing the selectivity of a search. In order to record and to use such characteristics conveniently and advantageously in accomplishing useful selecting and correlating operations, special programming of so-called general-purpose computers or the application of specially designed electronic selectors is necessary.

Previous discussion has emphasized the desirability and advantages of precision in the characterization of the subject contents of documents in order that searching and selecting operations may be performed with reliability both with regard to retrieving pertinent items and rejecting extraneous material. Such precision was shown to be particularly important when high degrees of reliability are required of searches defined in terms of logical products. Each term in the logical product must be aligned as precisely as possible with the scope of the information requirement. As already noted, the basic means for accomplishing such alignment is the definition of concepts and terms, together with care and precision in the use of terminology. These are, however, not the only means for accomplishing such alignment. A practical example may serve to point out another possibility. The "alkali metals" are usually defined in chemistry as the naturally occurring elements of the first group in the periodic table, namely, lithium, sodium, potassium, rubidium and caesium. An information request phrased as relating to some process that may make use of the metallic form of the alkali metals may pertain to all the above listed elements or, because of practical considerations of the high costs of the less common alkali metals, it may be desired to restrict the search to the process in question when performed with metallic sodium and potassium. In this case, the characteristic "alkali metals" would be too broad and might result in selection of extraneous materials. Instead of defining the search by using the term "alkali metals," an alternate definition "sodium and/or potassium" might be set up. Such a formulation is a particularly simple example of a logical sum, which might be symbolized in this case as $A + B$, where A means sodium, B potassium and the plus sign denotes "and/or". Logical sums of a larger number of terms may be searched with varying degrees of convenience with different types of equipment. With hand-sorted punched cards, the limited possibilities of independently recording characteristics imposes rather obvious restrictions. With aspect cards, as exemplified by the "Peek-a-boo" or "Uniterm" type, the inconvenience, not to say impracticality, of performing the multiple comparisons required also restrict the extent to which logical sums may be

employed in practice. With information systems using machine-sorted punched cards, searches defined with the aid of logical sums are readily feasible and, in some situations at least, such searches have proved highly advantageous. With fully automatic electronic selectors, the only limitation placed on the use of logical sums in defining searches is the availability of plugboard or similar capacity for programming the equipment.

The possibility of making use of logical sums to define the scope of a term has been exemplified in the preceding paragraph by directing attention to the term "alkali metal." The ability to vary the scope of a term by making use of logical sums is particularly valuable when adjusting the scope of meaning of terms in a logical product so as to insure high reliability of retrieval of pertinent information and simultaneous reliability of rejection of non-pertinent items. Such exploitation of logical sums requires, as is perhaps obvious, the possibility of formulating combinations on a multi-level basis as summarized in Chapter 2, pages 250-252.

In planning and conducting searches, it may be advantageous to specify that items to which a certain characteristic pertains are to be rejected. Specifications of absence of one—or more—characteristics is practically always coupled with the requirement that at least one characteristic shall be present. A simple example might specify that characteristic A shall be present but B shall be absent, and this might be symbolized as the logical difference $A - B$. Such negative specification of scope of search has not found, as might be expected, as extensive application as specification in terms of logical product and logical sum. As a consequence, it is much more probable that characteristics will be used as terms in logical products and it is, of course, desirable that such positive specification shall result in retrieval of all items of pertinent interest. On the other hand, when a characteristic is used, on occasion, in a logical difference to reject certain items, it is desirable that information of pertinent interest shall not be rejected. When used in logical products, it may be desirable for a characteristic to be applied to a somewhat wider range of documents than is the case when the same characteristic is used as a term in a logical difference to achieve rejection. Here again, the possibility of defining the characteristic in alternate fashion by means of logical sums can often be turned to advantage. In this way, the possibility of formulating logical sums enhances the usefulness of logical differences as a means for formulating the scope of searches to be performed by mechanical or electrical means.

Cost Considerations and Achievement of Benefits

Other chapters in this book and the preceding sections of this chapter have indicated a considerable range of devices and equipment that may be

applied for searching and selecting purposes. Correspondingly, methods of code construction range from simple direct coding (one of the most advantageous methods when used with hand-sorted punched cards) to encoded abstracts designed to be searched by fully automatic selectors. In deciding which coding method—and also which equipment—can be expected to be most advantageous in a given case, guidance can be provided by considerations of costs, on the one hand, and of benefits, on the other hand. The following underlying considerations are particularly important.

1. Nature of information requirements to be serviced. (In evaluating future information requirements it is, of course, short-sighted to take into consideration only previous patterns of use of libraries and files. To achieve maximum benefits from new searching and selecting methods, their imaginative evaluation is a virtual necessity.)

2. Informational material for which the system is being designed.

(i) Character of the information, especially, degree of complexity, extent to which it is fragmentary in character, extent to which it partakes of the nature of raw data and requires interpretation or mathematical processing.

(ii) Number of documents to be embraced.

(iii) Form of recording of information, e.g., printed or written documents, photographic or diagrammatic material, etc.

(iv) Probability of obsolescence of the information and rate of obsolescence.

3. Requirements imposed on the system.

(i) Selectivity—ability to limit the number of items to which personal attention must be directed.

(ii) Correlative capability—especially in utilizing incomplete or fragmentary information.

(iii) Reliability requirements—acceptable levels of probability of retrieving pertinent information and of probability of excluding extraneous material.

(iv) Promptness requirements—permissible time interval between formulation of an information requirement and its servicing.

(v) Form of supplying needed information, e.g., bibliography, collection of abstracts or of original papers.

4. Problems in analysis and characterization of information.

(i) Status of concept formulation in the field to which the information pertains.

(ii) Status of nomenclature and terminology—degree of general agreement as to both meaning of individual terms and also their interrelationships.

(iii) Codification problems, e.g., codes for organic structural formulas or electrical wiring diagrams.

(iv) Recording of observational relationships, e.g., mode of interaction between various substances or organisms, or mode of their involvement in various more or less well-understood processes.

With such considerations in mind, the principal sources of cost may be evaluated:

1. Acquisition and preliminary processing of information preparatory to its analysis and encoding for subsequent searching and retrieval.

2. Analysis of information as to important features. (More detailed analysis, if skillfully performed, can provide the basis for increased selectivity of the searching operations. On the other hand, more detailed analysis is virtually certain, other factors being the same, to incur greater costs.)

3. Encoding and recording the results of analysis. (With coding systems that provide for detailed analysis of subject matter, automatic encoding methods may enable considerable savings in costs to be effected. With recording media that do not permit automatic procedures to be applied, any increase in complexity of a code almost unavoidably means increased costs in this step.)

4. Cost of analyzing and programming information requirements. (With the simpler—but less selective—devices, such as hand-sorted punched cards, the limits within which these costs may vary are quite narrow. This is not the case, however, with complex electronic equipment capable of working with detailed code systems as exemplified by the encoded abstracts described in Chapter 11 and elsewhere¹⁶.)

5. Cost of conducting searching and selecting operations. (In minimizing this cost, proper choice of equipment is a particularly important factor. It is as inadvisable to use a highly complex electronic machine for a small file that can be well serviced with hand-sorted punched cards as it would be to attempt to use the latter for applications in which the necessity of meeting diverse information requirements by searching extensive files makes manual methods much less effective than automation procedures.)¹⁷

6. Cost of equipment. (The choice of optimum equipment requires that attention be devoted to such factors as selectivity requirements, degree of detail of characterization of information, required speed in servicing requests for information, form in which information must be furnished.)

¹⁶ J. H. Shera, Allen Kent and J. W. Perry, (eds.) "Information Resources. A Challenge to American Science and Industry," New York, Interscience Publishers, Inc., 1958.

¹⁷ See footnote 7, page 399.

Conclusion

In applying various devices and types of equipment to search, to select and to correlate recorded information the latter's characteristics as well as the scope of information requirements to be serviced must be expressed by a common code. Its design is of decisive importance both in meeting the requirements that are made of an information system and in keeping costs at the lowest possible levels. Other things being equal, the simplest code that is effective in meeting requirements is to be preferred. Any elements of complexity in a code must justify themselves by providing benefits commensurate with costs.

Chapter 19

HOLES, PUNCHES, NOTCHES, SLOTS, AND LOGIC

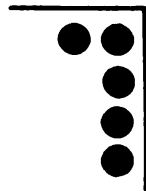
C. D. GULL*

National Academy of Sciences-National Research Council, Washington, D. C.

This chapter attempts to describe the fundamental significance of holes, punches, notches, and slots in punched cards in relation to their positions, assigned values, and use in the storage and retrieval of information. It will be seen that the information content of a punched card, ignoring the written information thereon, is a long expression in binary code, and that the words, numbers, and other symbols making up this expression can be subjected to certain mathematical and logical operations. The first edition of this book was four years old when R. A. Fairthorne, of the Royal Aircraft Establishment, Hants, England, pointed out that several of these elementary considerations had been unknown or overlooked in the description and mathematical analysis of punched cards and codes in the first edition¹.

Definition of Kinds of Holes

The various kinds of holes are defined in this chapter as follows:



- (1) *Guide holes*—one or more rows of holes pre-punched around the margins of hand-sorted cards. Although it is customary to think of these holes as having the values of the symbols printed alongside them, they have no value and their sole usefulness is to permit a sorting needle to be inserted anywhere in a pack of cards in a standard pattern of holes.



- (2) *Punches*—any kind of holes punched out of machine-sorted cards.



- (3) *Notches*—the bits punched out of hand-sorted cards to connect pre-punched guide holes on the outer row with the nearest outside edge; or the bits punched out from an inner guide hole but not connecting two holes.



- (4) *Slots*—the bits punched out of hand-sorted cards to connect two or more guide holes; slots can be extended to the outside edge with notches.

* In 1958, Mr. Gull joined the Computer Department of the General Electric Co.

Punches, notches, and slots are assigned value; their presence indicates that the analyzed document represented by the punched card possesses the attributes assigned to the areas punched.

Punched Cards and Binary Codes

The information content of a punched-card can be defined remarkably well in binary codes, as shown in the following statements, all of which can be expressed as "one or zero," or reduced to a binary expression:

- (1) A card offers two surfaces on which to record information.
- (2) Any position on either surface can be identified as the coordinate of the horizontal and vertical axes.
- (3) Any position on a card can be punched or unpunched.

In contrast to a hole, a mark on a card (text) has significance not only for the side and location where it is placed, but particularly for its configuration and relationship to adjoining marks. There is such a variety of configurations in the characters of the world's alphabets, number systems and script writings, and the sequence of characters is of such paramount importance, that the simpler values of side, position and condition (punch or no punch) are easily overlooked; but these are the determinants for punched cards in this chapter, in which the written text on cards is not further considered.

A Card has Two Surfaces, Zero or One. Customarily one corner of each card in a file is cut off to facilitate checking to see that all cards face in the same direction. When uniformity of facing is thus assured, each side of the hand-sorted card can be printed with a different code and distinguished by using one guide hole to indicate the facing. When only the front is used, the hole can be left unnotched, and it can be notched when the back is used. When both sides are printed in different codes, the notched and slotted pattern becomes a form of superimposed coding and unwanted cards or drops* may be expected. False drops can be isolated by reading the pattern of the notches from the other side and checking to see if this pattern has been assigned in the code on that side.

Any Position can be Expressed in Binary Code as the Coordinate of Two Axes. The guide holes of hand-sorted cards are always in the same positions, otherwise the sorting needle could not be inserted through a pack of cards. Machine-sorted cards are always punched very precisely on a gridiron pattern at the coordinate point of a horizontal row and a vertical column, else the existence of a punched hole can be missed when a coordinate point is tested on a pack of cards by a mechanical or electrical impulse. One common numbering system of hand-sorted cards is 1, 2, 3, $\dots$ n , reading from right to left and starting fresh with 1 at each corner; IBM

* For discussion of superimposed coding and false drops, see Chapter 21.

cards are commonly numbered 1, 2, 3, . . . 80 across the columns left to right, and (12, 11) 0, 1, 2, . . . 9 reading down the rows. Any position can be described as the coordinate of two arabic numbers and converted to the binary equivalents of the decimal digits.²

The Two States of Punched Cards and Their Holes. Except for hand-sorted cards which are pre-punched with guide holes, punched cards have only two states—punched and unpunched. This is also true of a single punching position, and single cards with their many punching positions customarily show a minority of positions punched out and a majority unpunched. Since the punched state can be considered 1 (one) and the unpunched state 0 (zero), it is clear that the face of each card is fundamentally an expression in binary code, regardless of the arbitrary values assigned to the punching positions. This code is generally employed to search for “ones,” but the “zeros” can be searched for with profit too.

Values Assigned to Punches, Notches and Slots

Letters, Numbers, and Other Notations. As described previously, information is recorded in hand-sorted and machine-sorted cards by punching out bits of the cards. In machine-sorted cards the value is assigned to the coordinate point and this is punched out, but in hand-sorted cards part of the card adjacent to and connecting with the pre-punched guide hole is notched or slotted out since the coordinate point was removed by the pre-punching operation. Thus, while punching and notching are mechanically the same for hand- and machine-sorted cards, the values are different for the resulting holes.

When a card is used, a number, at least, is punched or typed on it corresponding to the number on the original document; and a catalog entry plus an abstract is commonly recorded on hand-sorted cards. Until the document is analyzed for subject content, the pattern printed on the card has no value; all that can be said of any punching position before the document has been analyzed is that its value is “undetermined.” After analysis and punching, the notches and punches individually acquire the value “one” (or “yes”) and the unpunched positions are determined to be “0” (or “no”). In practice most of the possibilities for punching are never seriously considered and so most of the unpunched positions can be considered to have the value of “0” (zero).

The conditions “1” and “0” are read according to the values printed on each card for the punching positions. Thus on a hand-sorted card (Figure 19-1, right), the value of the unnotched holes is “not a, not b,” etc., and the value of the notch is “c”; on the machine-sorted card (Figure 19-1, left) the value of the punch is “4,” and “not 11, 12, 1,” etc.

A notch represents a shift from one coordinate to another; a punch repre-

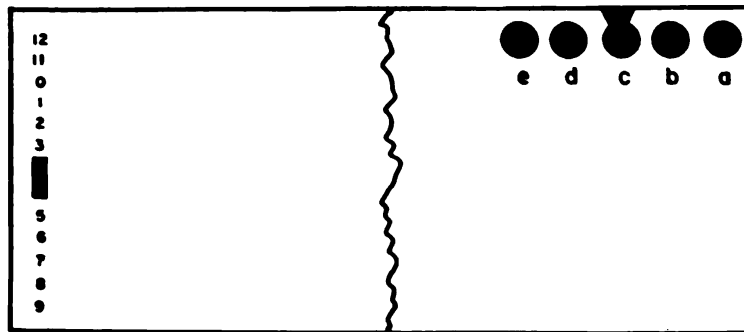


Figure 19-1. Representation of a punch in a machine-sorted card (left) and a notch in a hand-sorted card (right).

sents a change from no hole at a coordinate to a hole there. These differences reflect the various ways of searching and sorting cards. A needle run through a pack of hand-sorted cards tests one hole for a “yes” or “no” answer; the “yes” answers are determined when some of the cards shift to a new position, to an adjoining hole, or drop off the needle, while the “no” cards remain fixed on the needle. In machine-sorted cards, one card is tested after another; cards are not tested in packs. In testing a column of 12 coordinate points, a “yes” answer at any point directs the cards into the corresponding sorting pocket; a “no” answer from all 12 points directs the card into the “reject” pocket, or sends an electric impulse along one of 13 wires for the column being tested to actuate other sorting and printing mechanisms.

In Chapter 2, Casey distinguishes between direct coding and indirect or combination coding. For direct codes the holes are assigned one value each; in combination codes the holes have shifting values, depending on which holes are notched in combination. Thus “1” and “7” notched in the units field of a numerical code signify “8”; in an alphabetic code they may signify “H”. Since two and three rows of holes around the edges of a card are in fairly common use, since slots can be punched out between holes, and since two or more values can be assigned to the holes of a field or overlapping fields by superimposed coding, it is obvious that the possible mathematical combinations are very numerous indeed. Wise treats the mathematical analysis of punched-card coding in Chapter 2 of this book, including a discussion of the probability of getting false drops or unwanted cards when superimposed codes are used.

The introduction of a notch or a slot between two holes greatly complicates the values assigned to the bits punched out. It is necessary to describe the logical operations which apply to the punching and searching operations before consideration of the values assigned to the punches can be completed. It is sufficient to conclude this discussion by saying that

any letter, number, symbol, or pictograph can be assigned to any punching position, and that the value of the bit punched out will be the result of one or more of the following factors:

- (1) The value assigned to the punching position.
- (2) The angular direction and distance of the shift away from the punching position.
- (3) The relationships created by the existence of other punches on the same card.
- (4) The selection and sequence of the searching operations (as reviewed in the following paragraph).

In Chapter 2 Casey discusses the elementary manipulations of hand-sorted cards after they are punched, and these may be described as follows:

- (1) Single pass searching—searching the whole deck of cards *once* with one or more needles.
- (2) Sequential sorting—the arrangement of a file of cards into an alphabetical, numerical or classified order, or into one major order with subarrangements in another order, by a succession of single passes of one or more needles through the whole deck for each pass.
- (3) Sequential searching—a single pass through the whole deck with one or more needles, followed by a succession of single pass searches performed on the ever diminishing pack of selected cards.

Logic and Punches, Notches and Slots

A file of punched cards, the punches in the cards, and the punching and searching operations can be described in certain of the postulates and theorems of symbolic logic or modern formal logic, and shown in the circles known as Venn diagrams.³ In Figure 19-2 the card is considered unity (or 1) and the area around it as its complement or zero. After the card is punched, it represents a document; the assemblage of all the cards in one file (that is, their logical conjunction or alternation) is also unity. The analysis of a

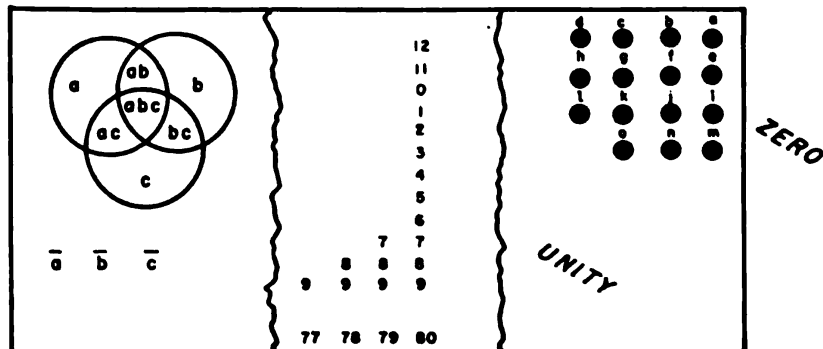


Figure 19-2.

document by words, numbers or symbols creates classes which are shown in the intersecting rings labelled a , b and c . The class letters and numbers are shown on the right to suggest Keysort and IBM cards. The total number of classes is limited only by the punching positions available and by the number of classes which can be generated by the physical, mathematical and logical combinations of these positions. Any number of circles or irregular figures can be drawn to intersect as desired, but three circles are sufficient for demonstration here. Thus the intersection or conjunction of three figures is written abc , of four $abcd$, of five $abcde$, etc. The area outside any figure is its complement; thus a is also not b and not c (written $\bar{b}$ and $\bar{c}$) and ac is also $\bar{b}$ and ab is also $\bar{c}$. Assume for the next few paragraphs that unity is limited to three classes, a , b , and c .

When class a is found in a document, a is notched out of a hand-sorted card and punched out of a machine-sorted card to record this finding. Assuming that classes b and c are not found in the document, positions b and c are undisturbed; and the document can be shown as $a\bar{b}\bar{c}$.

When class a is the object of a search, the cards are tested in position a . The a 's fall off the needle of hand-sorted cards and the others remain on the needle. The presence of a is shown by the translational movement of the card; its absence by a failure to move. When a is tested on a machine sorter, cards with a hole in position a are directed to the corresponding pocket, and cards not punched in position a are sent to the $\bar{a}$ (or reject) pocket. The test of machine-sorted cards is for the presence or absence of a hole; the test of hand-sorted cards is for the presence or absence of movement of the cards.

Assume that class b is found in the same document. Figure 19-2 now appears changed as the cards are customarily used (see Figure 19-3). In the algebra of logic this condition can be written $(a + b)\bar{c}$, and spoken of as " a and b conjoined with not c ." $(a + b)$ is the logical sum, also called an alternation, for it means "either a or b , or both a and b ."

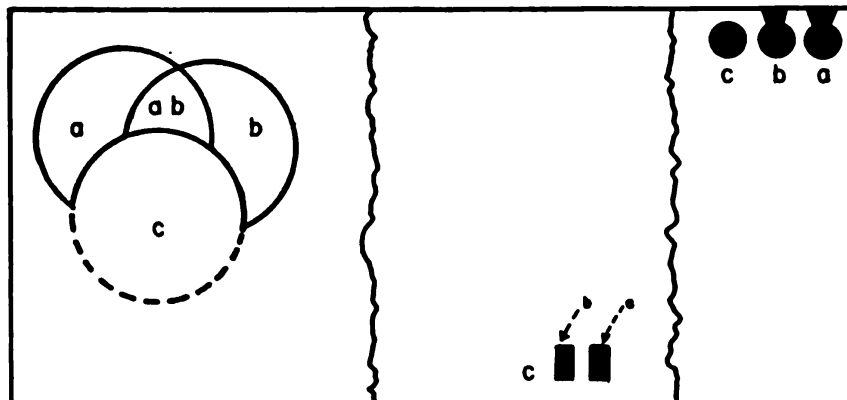


Figure 19-3.

Now the cards can be searched for a and b , using two needles or two brushes in the proper positions, in a single pass search. As shown in the Venn diagram, the class ab is possible here. It is the logical product or conjunction of classes a and b (written $a \times b$, $a \cdot b$ and ab , and by the law of commutation, ba). Unless the document is carefully analyzed, the possible presence of ab may be overlooked. Whether the conjunction ab is overlooked or not, a single pass search for a and b actually yields $a + b + ab$. Unless a sequential search is then made for $\bar{c}$, any of the cards found to $a + b + ab$ may contain c in any of its forms, ac , bc , and abc ; that is, the documents may treat subjects beyond those sought, unless all positions are searched.

Notches can be used in hand-sorted cards to distinguish between alternations and conjunctions in certain circumstances, but machine-sorted cards as they are now used do not offer this attribute. When a slot is punched between a and b , hand-sorted cards can be needled from a to b or b to a , in searching for $ab = ba$; and a , b , or $a + b$ cannot be obtained, for they cannot drop off the needles. Also, when a and b are notched, the cards cannot be moved horizontally and consequently ab cannot be obtained. Since there are no brushes to search the area between a and b on machine-sorted cards, this equipment cannot distinguish in a single inspection between $(a + b)$ and ab . A punch can be made elsewhere in these columns with the value "conjunction—yes," which will serve to distinguish ab from $a + b$ (when only one class is punched out of each column).



Assume that class c is found in the same document; the Venn diagram of Figure 19-2 now applies. There are three punches, a , b , and c in the machine-sorted cards, and ab , ac , bc and abc cannot be distinguished from $a + b + c$. Three notches in the hand-sorted cards mean $a + b + c$; a slot from a through b to c means abc ; a notch in a and a slot from b to c mean $a + bc$, etc. There are 5040 permutations of the seven positive classes (a , b , c , ab , ac , bc and abc) which could be punched into the card for these three positions, but fortunately for ease in analyzing documents, most of the permutations have no practical value; and certain of the card movements create overlapping classes, while others are physically impractical. Thus the slot from a through b to c , which is abc , means for example that ab cannot be distinguished from $ab + bc + ac$, nor from abc . A physical barrier is needed at b to distinguish ab from bc , and the slot for ac needs to be curved away from b for it to become distinctive.

A more complex situation exists in hand-sorted cards if we consider posi-

tions f and the eight positions a, b, c, e, g, i, j and k immediately surrounding it. It has been conventional practice by some to slot f to b to record f (called intermediate punching), and by others to slot f to b and through to the outer edge of the card to record f (called deep punching). The previous paragraphs have shown that bf and $b + bf$ are actually punched in these operations. The punch above position f should extend only halfway to position b to enable f to be recorded and searched specifically. The close proximity of the pre-punched holes makes it difficult and nearly impossible on some hand-sorted cards to notch a single class in an inner row of holes without connecting two holes.

Slots can also be punched from f to radiate to any adjacent hole, creating af, bf , etc., and in various patterns such as the right angle formed by fei . It is interesting to note here that fie is not the same pattern as fei , although $fei = ief$ and $fie = eif$ (Figure 19-2).

A very complex situation exists around f without slotting through to the eight adjacent guide holes. Assuming that holes, punching devices and needles are properly balanced in size, eight notches can be punched out, all radiating from f at 45° intervals. By rotating a pack of cards around f , cards can be made to drop the distance of a notch for any of the eight directions. It is now seen that the value of a guide hole is determined by two immediate factors, the practical number of directions the card can be shifted away from it and the distance of the shifting motions, as well as by the mathematical and logical combinations already considered for two or more notches and slots.*

Application of this Analysis

Assignment of Values to Punching Positions. Four methods of arrangement can be used in assigning values to punching positions on punched cards:

- (1) Random arrangement of letters, numbers, symbols, words and phrases, assigned from time to time whenever they are needed to a pre-printed standard card.
- (2) Serial arrangement of numbers, or repetitive patterns of number sequences.
- (3) Alphabetic arrangement of letters, single words, phrases, and subject headings.
- (4) Classified arrangement of letters, numbers, words and symbols. It is customary to maintain auxiliary indexes and classification schedules to insure effective use of the values printed on the punched cards. Various

* It is quite probable that mathematical set theory can be applied with profit to extend the analysis of this section. Lack of both time and background has prevented me from undertaking to examine the subject of this chapter with set theory.—CDG.

expedients are used in the auxiliary lists and on the cards to maximize the capacity of the punching positions and minimize the retrieval of unwanted cards. Some of the expedients in use are only partially effective in accomplishing this objective.

It is essential that the designer understand and apply the analysis of the previous section in order to obtain the best results from the design to be printed on the cards. The most difficult task is the design of a punched card for use with a classification, and one example is given to show the economy and precision which can be obtained with careful analysis of the assignment of punching positions.

The second subdivision of the "ASM-SLA Metallurgical Literature Classification"⁶ is the Materials Index; it is further subdivided into four parts:

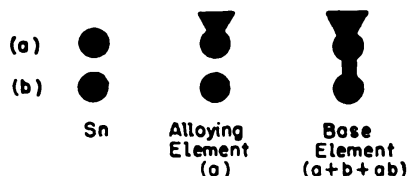
- Common Elements and Their Alloys
- Other Elements and Their Alloys
- Properties and Applications
- Ferrous Groups.

The Common Elements are arranged alphabetically by chemical symbol:

- Ag—Silver and its alloys
- Al—Aluminum and its alloys
- (...and on to)
- Zn—Zinc and its alloys.

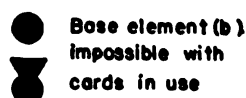
In the card designed for this classification two holes are assigned for each element. A notch from the outer hole to the edge of the card indicates that the metal is the alloying element (less than 50 per cent of the alloy) and the slot plus notch from the inner hole to the outer edge (the "deep punch") indicates that the element is the base metal (more than 50 per cent of the alloy). Provision is also made in another 32 holes for further subdivisions, such as the unalloyed (or pure) metal, a metal's alloys in general, alloys by name (such as brass and bronze under copper, and solder under tin), and for more specific alloys under these. The holes have shifting values dependent on the elements; the 16 inner holes are labeled *a* through *s*, and the outer 16 holes are labeled 30 through 45. Thus "white metal" is "Sn-d-32," requiring four holes since "Sn" is the "deep punch" here.

The analysis of Logic and Punches, Notches and Slots (pp. 426–8) was applied to the original use of two holes and "shallow and deep punching" for the alloying and base metal conditions, which were:

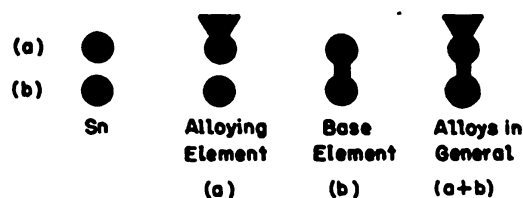


and it was seen that the "deep punch" did not retrieve the base element alone when needed, but retrieved Alloying, Base, and Alloying X Base, or in symbols, $a + b + ab$.*

Since the space between holes on the cards in use is too small to permit notching of inner holes,



it was recommended that this distinctive arrangement be adopted:



Since another hole (an inner hole labeled b) had been required to indicate alloys in general, this recommendation had the merit of freeing that hole for another assignment. It also permitted the established assignment to be continued in existing files, and it prevented false drops occasioned by the retrieval of a base metal when an alloying metal was sought.

References

1. Fairthorne, R. A., "The Patterns of Retrieval," *Am. Doc.*, 7, 65-70 (April, 1956).
2. Holden, W. H. T., Digital Codes for Numerical Control. *Electrical Manufacturing* 58, 105-11 (Dec. 1956).
3. Taube, Mortimer, and Wachtel, Irma S., "The Logical Structure of Coordinate Indexing," Chap. 8, pp. 73-76, of *Studies in Coordinate Indexing* (V. 1) by Taube and Associates, Washington, Documentation Inc., 1953.
4. "Class Definition and Code Construction," Chap. 6, pp. 27-33, of *Machine Literature Searching*, by James W. Perry, Allen Kent and Madeline M. Berry, New York, Interscience, 1956.
5. "ASM-SLA Metallurgical Literature Classification," prepared by a Joint Committee of the American Society for Metals and the Special Libraries Association, Cleveland, ASM, 1950.

* The conjunction ab here is a representation of the alloying x base condition in equilibrium diagrams of alloys, provided elsewhere in this classification.

Chapter 20

CORRELATION OF RESEARCH DATA AND ESTABLISHMENT OF CAUSE AND EFFECT RELATIONSHIPS

D. E. H. FREAR

The Pennsylvania State University, State College, Pa.

Introduction

Correlation is the process of relating one set of measurements to another. For example, it is not difficult to obtain the weights and heights of a group of experimental animals. These data by themselves bear no relation to each other, but it is frequently important to know to what degree an increase in height is associated with an increase in weight. There are many problems in scientific research where a study of relationship or association is of greater importance than an investigation of the variation of any one factor. Sometimes—although not always—it is possible to establish that variations in one factor are the cause of corresponding positive or negative variations in another.

The approximate measure of the relationship existing between two factors may be estimated for some types of data by plotting the variables graphically. If the resulting scatter diagram or graph shows a definite trend in one direction, it can be assumed that there is a correlation between the two measurements.

Scatter diagrams give only qualitative information, and for most cases it is desirable to be able to express the relationship by a definite numerical value so that one lot of data may be compared to another. To accomplish this, the *correlation coefficient* is usually employed. This value is calculated

$$r = \frac{\Sigma xy}{N(\sigma_x \sigma_y)}$$

from the formula in which r is the correlation coefficient, x and y the deviations of the x and y values from their respective means, σ_x and σ_y the standard deviations of the two characters, and N the number of pairs of characters. The standard deviations are calculated by taking the square root of the sum of the squares of all deviations divided by the degrees of freedom. The correlation coefficient may vary from $+1.0$ to -1.0 , and may be positive or negative. The significance of any correlation coefficient

depends on its relation to its standard error, which should always be calculated^{1, 2}.

Punched Cards in Correlation Studies

Although with small groups of data it is relatively easy to determine correlations qualitatively by plotting the information graphically, or quantitatively by calculating correlation coefficients, the labor involved when dealing with large numbers of individual values becomes formidable. Since in many cases the use of large statistical populations increases the accuracy of the resulting information, it is obvious that the more accurate the results one wishes to obtain the more work is involved. It has been stated³ that in a problem having 492 observations 2,952 multiplications are necessary.

The use of punched cards markedly reduces the work involved in correlation measurements, particularly where large numbers of individual figures are to be studied. It should be made clear that the use of punched cards reduces the time and labor in many types of calculations; in general, the same information may be obtained by conventional methods if time and labor costs are not important factors. Some of the ways in which punched cards may be used to advantage are given below. Since in a limited space it is possible to give only a few examples, an attempt will be made to make the illustrations as general as possible. It will be necessary to work out individually further adaptations to specific problems. The reader will find that the manufacturers of punched-card equipment are always ready to suggest ways and means of using their products, and it is suggested that questions on unusual applications be directed to them.

Qualitative measurements of correlation. As mentioned earlier, the simplest type of correlation studies are those of a qualitative nature. Where only a few measurements are involved, a graph serves the purpose quite well. Punched cards of the edge-punched type may be also used directly for this purpose. Clarke⁴ has pointed out that this type of card will serve as a three-dimensional graph if the notches punched on the four edges are arranged in numerical order. Suppose, for example, that we have several thousand samples of wood on which physical and chemical measurements have been made. For each sample a card is punched, giving specific gravity on edge A and tensile strength on edge B. By simple sorting, these cards

¹ Ezekiel, M., "Methods of Correlation Analysis," 2nd. Ed., New York, John Wiley and Sons, 1941.

² Paterson, D. D., "Statistical Technique in Agricultural Research," New York, McGraw Hill Book Co., 1939.

³ Brandt, A. E., *J. Am. Statistical Assoc.*, 23, 291-5 (1928).

⁴ Clarke, S. H., *Nature*, 137, 535 (1936).

may be arranged in increasing order of specific gravity, thus giving a regular ascending progression of punched notches along edge A. A visual observation of edge B will suffice to tell whether the notches on this edge fall into a regular positive or negative progression, or are randomly distributed. A third and fourth measurement can be added to the cards in the same way, and any two or more of the measured factors correlated by suitable sorting.

This simple method of detecting obvious relationships between two or more variables appears to have considerable use in studies where only qualitative results are required, or in cases where it is difficult to decide whether or not the labor involved in a more complicated mathematical determination of the correlation coefficient is justified. The punched cards prepared for this study generally have other purposes, such as classification or identification, so that the data on correlation obtained in this way may be considered a very useful by-product.

When large numbers of individual measurements are involved, punched cards can give qualitative and quantitative information on correlation in another way, by making it possible to group the data into units which can be handled easily. For example, let us suppose that 10,000 samples of soil have been analyzed for calcium and magnesium. If the data are punched on cards, it is easy to separate the cards into 50 groups of increasing calcium content, for example, from 0.10 to 0.11 per cent, from 0.11 to 0.12 per cent, etc. Each group may contain several hundred cards; the groups may be made larger or smaller by regulating the limits. The mean magnesium content of each group of cards may then be calculated; instead of 10,000 individual figures, only 50 remain, and the degree of correlation may be determined by inspection, or by calculation, if an accurate numerical value is desired. This method of treating data is discussed in some detail by Carver⁵.

Semiquantitative correlation studies. Into this category fall a number of statistical procedures which yield information on the relationship between two sets of values, but do not give correlation coefficients as such. There are many means of securing information of this type, and the actual methods to be used will depend on the problem at hand. One example will illustrate the general method of attack.

In a study of a large number of chemical compounds of known structure it was desirable to know if certain chemical groups were associated with toxicity to insects. Inasmuch as there were over 6,000 different compounds involved, each of which contained an average of over four distinct chem-

⁵ Carver, H. C., "Uses of the Automatic Multiplying Punch," *The Punched-Card Method in Colleges and Universities*, Baehne, G. W., Ed., 421-2, New York, Columbia Univ. Press, 1935.

ical groups, the problem was obviously too large to solve by conventional methods.

The first step in the solution entailed the preparation of a key list in which the chemical groups were assigned numerical designations⁶. Punched cards, each having a number punched corresponding to the chemical group present, were designed for each compound. Thus for ethyl alcohol, with an ethyl and a hydroxyl group, two cards were punched, one bearing the numerical designation of the ethyl group, the other for the hydroxyl group. In another location on each card the toxicity of the parent compound was indicated by punching.

When all of the cards were prepared, it was easy to separate all compounds having a hydroxyl group, for example. These were then sorted for toxicity, and it was found that 82.8 per cent of the insecticidal tests made with these compounds were positive. When all of the chemical groups present in the compounds were thus analyzed, some rather striking differences appeared. It was found, for instance, that compounds containing the thiocyanate group were consistently high in toxicity, while those containing amino groups were low⁷.

This specific example, while somewhat out of the ordinary, is nevertheless typical of a number of correlation problems. Punched cards are particularly useful in the solution of such problems when large numbers of individual figures are involved. By preparing a key, almost any property may be translated into punches on the card. By sorting, either by hand or mechanically, selected properties may be related to one another, or the frequency of occurrence of one happening may be related to the occurrence of another.

Quantitative correlation studies. For mechanically sorted punched cards there is a wealth of information available on methods and procedures for making correlation calculations^{8, 9}. By using the proper machines and with the correct wiring of the plugboard it is possible to obtain answers to relatively complex calculations. The automatic multiplying punch is required for this purpose, and the machine may be so wired that it will record the product xy for any set of data, as well as Σxy . By double-wiring the machine the Σx^2 and Σy^2 values may also be recorded so that correlation studies on large populations may be made in a short time. Tabulators may be used in similar ways.

⁶ Frear, D. E. H., "Catalogue of Insecticides and Fungicides," Waltham, Mass., The Chronica Botanica Co., 1947.

⁷ Frear, D. E. H., and Seiferle, E. J., *J. Econ. Entomol.*, **40**, 736-41 (1947).

⁸ Anonymous, "The Mendenhall-Warren-Hollerith Correlation Method," Document No. 1, New York, Columbia Univ. Statistical Bureau.

⁹ Eckert, W. J., "Punched-Card Methods in Scientific Computation," New York, The Thomas J. Watson Astronomical Computing Bureau, Columbia Univ., 1940.

In this application the punched-card machines become a type of calculating machine, as distinguished from their normal, more widely-known sorting functions. It is hardly possible to suggest a mathematical calculation which cannot be carried out by punched cards; complex spectra¹⁰, autopsies¹¹, and other medical data¹² have been treated in some way with these techniques. Multiple correlations and analyses of variance and covariance may be done on punched-card tabulators, following the methods of Brandt¹³, while other applications of punched cards have been developed at Iowa State College¹⁴. The field is so complex and the applications so specific that it is advisable to treat each problem individually and to secure the aid of a competent expert in the field.

Evaluating Correlation Results

Whether information is obtained by conventional mathematical methods or by use of punched cards, the same generalizations may be made regarding their interpretation¹.

Broadly speaking, correlation coefficients tell nothing of the reason for the relation between the correlated factors. In some cases there may be nothing more than a fortuitous relationship; in others, the relationship may be due to one or more common causes, or one of the measured factors may be acting to produce changes in the other. The last instance is, of course, the only case of true cause and effect relationship.

It is usually, but not always, possible to determine into what class a given correlation coefficient falls. For example, if one correlates the height of potato plants grown under varied conditions with their yields, it is possible that a high degree of correlation may be found. It is reasonable to assume that variations in both height and yield are the result of the same climatic and other factors. In this instance this conclusion would be more tenable than to assume that height influenced yield, or vice versa. If, however, the plants were grown under identical environmental conditions, it would be quite logical to infer that tall plants produce more potatoes than short ones.

In all correlation studies the worker should carefully examine every possibility before drawing conclusions. The distinction between a true

¹⁰ Astanasoff, J. V., and Brandt, A. E., *J. Opt. Soc. Am.*, **26**, 83-8 (1936).

¹¹ Black-Schaffer, B., and Rosahn, P. D., *Yale J. Biol. Med.*, **15**, 575-86 (1942).

¹² Edwards, T. I., "Public Health Repts.," *U.S.P.H.S.*, **57**, No. 1, 7-20 (1942).

¹³ Brandt, A. E., "Uses of the Progressive Digit Method," *The Punched Card Method in Colleges and Universities*, Baehne, G. W., Ed., 423-36, New York, Columbia Univ. Press, 1935.

¹⁴ Homeyer, P. G., Clem, M. A., and Federer, W. T., *Research Bull.*, **347**, Iowa Agr. Exp. Sta. (1947).

cause-and-effect relationship and one due to a common cause is frequently difficult to make, and an unsuspecting analyst may find himself placed in an untenable position because some obvious factor has been overlooked. Under such circumstances it is best to be as conservative as possible; present the data and give all pertinent information on how the calculations were made, as well as a complete discussion of the variables—whether measured or not—in the experiment. Draw conclusions only if they are clearly indicated, otherwise, present the data with suggested explanations only.

Chapter 21

MATHEMATICAL ANALYSIS OF CODING SYSTEMS*

CARL S. WISE

Northern Regional Research Laboratory, Peoria, Illinois

Introduction

In considering how best to code punched cards we soon become aware of a dilemma. The number of holes in the card is of necessity limited while the number of ideas that we may wish to code is very much larger. For this reason coding schemes developed to date are either (1) limited in the number of ideas that can be recorded or (2) give extra cards as a result of sorting operations. By the term "extra card" we mean one which responds to the mechanics of sorting but which does not contain information pertinent to the search being made.

An extra card may appear for probability reasons in the superimposed type of coding, as described in Chapter 10. Extra cards may also be obtained when other types of coding are used. Thus, for example, direct coding, as a result of an insufficient number of holes to code a very large multiplicity of ideas, may be forced to conduct the analysis of subject matter in terms so broad as to preclude supplying a precise answer to a given question. For example, a card file based on direct coding and concerned with carbohydrate chemistry might be able to indicate the use of ion-exchange resins in refining operations only by punching a hole marked "Refining—Miscellaneous Reagents". A mechanical sort directed to the hole designating such a heading may be expected to turn up cards not concerned with the use of ion-exchange resins in refining processes. These cards dealing with other special processing methods would be, of course, extra cards if the scope of our interest were limited to use of ion-exchange resins in refining processes. This point appears worthy of specific attention because it is sometimes assumed that only the superimposed coding systems can yield extra cards during sorting operations.

Because all types of coding developed to date have limitations in one

* This chapter is based on a paper entitled, "Chemical Literature Studies. Some Mathematical Possibilities in Mechanical Indexing and Sorting Systems", presented before the Division of Chemical Literature of the American Chemical Society at the 115th national meeting in San Francisco, Calif., April, 1949. Thanks are due the American Chemical Society for permission to publish this paper in revised form.

direction or another, an analysis of how to reduce these unfavorable factors to a minimum should prove helpful. By putting this analysis on a mathematical basis it becomes possible to include in the discussion not only (1) the capacity of a code in terms of number of subject entries that can be accommodated and (2) the probable frequency of appearance of extra cards, but also (3) the number of punches required for each coded entry and (4) various factors influencing the labor involved in using the file (e.g., the number of times the cards must be handled in order to conduct a search or in order to place them in a certain desired sequence).

Further discussion will be directed to three basic types of coding, namely, (1) direct codes, (2) selector codes, and (3) sequence codes. Superimposed coding systems, an example of which was described in Chapter 10, are also analyzed mathematically. It would be awkward, if not impossible, to present an analysis of coding procedures without referring to some specific type of card. Hand-sorted, edge-punched cards have been used as the basis for subsequent discussion in this chapter. Much—if indeed not all—of the reasoning developed in this chapter can be applied to machine-sorted cards of the “IBM” (“Hollerith”) and the “Remington-Rand” (“Powers”) systems. Other types of literature searching machines, such as the Rapid Selector, have also involved this type of reasoning^{8,9}.

Direct Codes

A direct code, by definition, uses a single hole to represent a single idea. In this system, if we have H different ideas to code, we will require H holes in each card. Or, to state the matter in a different way, if our card has H holes, direct coding will force us to analyze our subject matter in terms of not more than H concepts. The care and skill with which these concepts are selected are of paramount importance in setting up direct codes.

Direct coding cannot cause the sorting operation to produce extra cards from a mechanical point of view. Cards selected by needling a given hole have been coded for a given concept—no more and no less. As noted above, the fact that the analysis of subject matter is limited to H concepts at most may cause unwanted cards to appear when the file is consulted to obtain an answer to a specific question. Since skill in selecting concepts and terminology for subject-matter analysis cannot be evaluated or predicted mathematically, there is nothing more we can say concerning extra cards resulting from this type of coding.

Within the limits imposed by the restricted number of concepts in terms of which subject matter can be analyzed when using direct coding, it affords the simplest and most convenient method for conducting sorting operations. To sort on a given code entry (e.g., “Refining—Miscellaneous Reagents”) the cards only need be handled once and a single hole only need be needled.

With regard to arranging into a given sequence direct coding does not afford the maximum in ease and convenience. Consider, for example, the operations required to arrange in sequence a file of cards in which the numerals 0–9, inclusive, are coded directly in a field of ten holes. The mechanical manipulations necessary to arrange the file in sequence will consist of nine successive sorting operations. Since each sorting operation removes one-tenth of the cards from the pack (assuming uniform distribution), the total number of cards to be handled in arranging a file of 1000 will be $1000 + 900 + 800 + 700 + 600 + 500 + 400 + 300 + 200$, or a total of 5400 as pointed out by Keckley¹. A numerical sequence code, such as that described in Chapter 2, page 19, permits the cards to be arranged in numerical sequence with four sorting operations in which the entire file is handled each time. For a file of 1000 cards it is necessary to handle 4000 cards.

Selector Codes

Direct coding, as we have seen, attaches meaning to a single hole. It is also possible to attribute significance to a combination of holes. When this is done in such a way as to minimize the amount of mechanical sorting to isolate cards characterized by some one combination of holes in a given field, the resulting coding scheme is termed a "selector code".

The underlying mathematical law states that if C denotes the number of combinations of H things taken Y at a time, then

$$C = \frac{H!}{Y!(H - Y)!} \quad (\text{Equation I})$$

For example, if we attach meaning to a combination of two holes in a field of five holes, then we can indicate any one of ten concepts in that field, as we find by substituting in Equation I

$$C = \frac{5!}{2!(5 - 2)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{(1 \cdot 2)(1 \cdot 2 \cdot 3)} = 10$$

If the value of Y in Equation I were unity, then the code would revert to the direct type. From a mathematical point of view a direct code can be regarded as the limiting case in a series of selective codes.

With six holes the number of combinations available if two holes are punched is 15, as becomes evident by substituting in Equation I

$$C = \frac{6!}{2!(6 - 2)!} = 15$$

¹ Keckley, C. P., "Systems Service Department", New York, Charles R. Hadley Co. Private communication.

TABLE 21-1. NUMERICAL VALUES OF EQUATION I

		Selector Code Combinations											
		Y											
		1	2	3	4	5	6	7	8	9	10	11	12
1		1											
2		2	1										
3		3	3	1									
4		4	6	4	1								
5		5	10	10	5	1							
6		6	15	20	15	6	1						
H 7		7	21	35	35	21	7	1					
8		8	28	56	70	56	28	8	1				
9		9	36	84	126	126	84	36	9	1			
10		10	45	120	210	252	210	120	45	10	1		
11		11	55	165	330	462	462	330	165	55	11	1	
12		12	66	220	495	792	924	792	495	220	66	12	1

The maximum number of combinations is obtained when Y is equal to one-half of H if H is even or one-half of $H - 1$ or $H + 1$ if H is odd.

$$C = \text{Max. when } Y = H/2 \quad (\text{Equation II})$$

Thus, with a fixed number of holes punched in a field of six holes the maximum number of combinations is 20.

$$C = \frac{6!}{3!(6-3)!} = 20$$

In order to facilitate the use of Equation I in computing coding possibilities binomial coefficients are tabulated in Table 21-1. The numerical value of C can be read in the body of the table corresponding to the values of H and Y at the left and top, respectively.

In actual use selector codes may assume a variety of forms. Thus, codes are sometimes used in which all the combinations available are not employed for practical reasons. From the mathematical point of view this means that Y in Equation I assumes smaller values than is the case when it remains constant. For example, one of the common selective codes often used is 7, 4, 2, 1, 0, SF. Two needles are required for 1 to 9 inclusive (e.g., for 4, needles are inserted in the holes 4 and SF, and for 6, needles are inserted in holes 4 and 2). One needle only, however, is used for 0. The effect is somewhat wasteful with regard to coding possibilities for, as we have already seen, only five holes are needed to indicate ten numbers. Cox, Casey, and Bailey² have used a triangular five-hole field, and Wise³ has

² Cox, G. J., R. S. Casey, and C. F. Bailey, *J. Chem. Ed.*, **24**, 65-70 (1947).

³ Wise, C. S., "112th Meeting", New York, A. C. S. (1947).

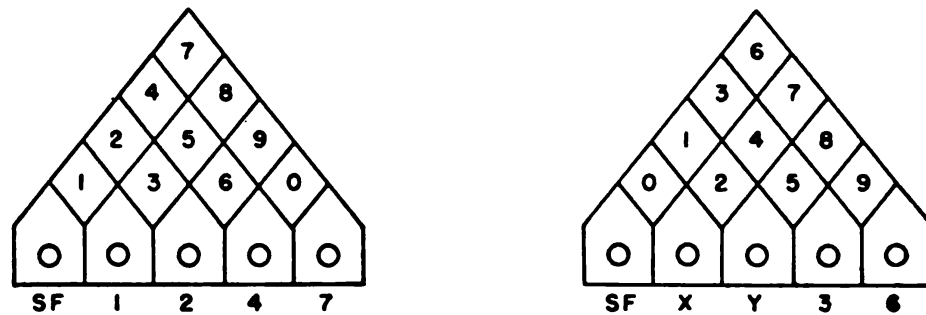


Figure 21. Alternate arrangements of numerical selector codes. (The single row of symbols may be used instead of the triangular arrangement of symbols above the holes).

used a non-triangular arrangement to indicate these ten numerals. These two pairs of arrangements are shown in juxtaposition in Figure 21-1.

Selector codes have been developed for letters of the alphabet. Thus, the author has described earlier^{3a} an eight-hole, two-needle selector code in which the holes on the card are labeled SF, A, B, D, G, K, P, V. To code A, the SF and A holes are punched, to code C, the A and B holes, while F requires punching of D and B, etc. The same holes may be used for numbers up to 28, in which case the holes are labeled SF, 1, 2, 4, 7, 11, 16, 22.

In order to evaluate the usefulness of selector codes it must be kept in mind that they permit only one numeral, letter, or—in the general case—concept to be coded in each field. Thus, in a field of five holes in Figure 21-1 only one numeral (letter or corresponding concept) can be coded. The reason for this is that coding a second numeral would cut so many notches in the card that it would drop out as an extra card with excessive frequency. For example, coding both 2 and 9 in the SF (2 + 7) triangular arrangement would cause the card to be selected as an unwanted card when needling positions 7 and SF to search for 7. Matters would be even worse if the numbers punched simultaneously were 2 and 8. A card so punched will appear as an extra card when searching for the numerals 9 (2 + 7), 7 (7 + SF), 3 (2 + 1), and 1 (1 + SF). Such a plethora of extra cards would prove intolerable in practice. These considerations led to the conclusion that selector codes are principally useful for coding a single member of a group of mutually exclusive concepts, of which the year of publication of a paper is a good example. Other examples are the first letter of a surname and the country in which some person or thing is located at a given time.

Our consideration of selector codes would be incomplete without pointing out certain factors which control the amount of time and effort needed to do the punching required for establishing the file and to conduct searches.

^{3a} Wise, C. S., "A Punched-Card File Based on Word Coding," in Casey, R. S., and J. W. Perry, "Punched Cards," 1st edition, New York, 1951, Chapter 6.

In this connection it is helpful to introduce the concept C_n , defined as number of combinations per needle, used in sorting. As mentioned above, punching two holes in a field of six gives 15 combinations, while punching three holes in the same field gives 20 combinations. With two positions punched two needles are required for sorting, the number of combinations is 15, and C_n is $15/2$ or $7\frac{1}{2}$; while with three holes punched C_n is $20/3$ or $6\frac{2}{3}$. The efficiency of the mechanical sorting is measured by C_n , and consequently it is desirable, other things being equal, to punch two rather than three positions in a field of six positions. We shall prove subsequently that, in the general case, C_n is a maximum when 37 per cent of the holes in a field are punched. These considerations as to the advantages of punching two holes rather than three in a field of six holes may be outweighed, in some cases, by the need for having 20 combinations available in the field rather than 15.

Selector codes can also be used for arranging in sequence. Like the direct codes, the number of passes of the needle required is $H - 1$ (i.e., one less than the number of holes in the field to which sorting operations are directed). Since fewer holes are required for the same number of coding possibilities using a selector code, cards so coded are in general arranged in sequence with less effort than is the case with a direct coded file.

Sequence Codes

Sequence codes, like the selector codes, are based on the principle of attributing meaning to a combination of holes. As the name implies, sequence codes are set up in such a way that the cards may be sorted into a predetermined sequence with a minimum of effort.

In discussing selector codes we concerned ourselves with combinations generated by taking a fixed number of holes Y in a field of H holes. Sequence codes are based on taking all possible combinations by permitting Y to vary from zero to H . Sequence codes make available for subject-matter analysis a number of concepts equal to the sum (C_s) of a series of selector codes (C); or, in mathematical language, from which $C_s = 2^H$.

$$C_s = \sum_{Y=0}^{Y=H} \frac{H!}{Y!(H-Y)!} \quad (\text{Equation III})$$

Equation III is, obviously, a summation equation based on Equation I.

A sequence code offers many more coding combinations than a selector code. Thus, in a six-hole field we have, by substituting in Equation III

$$C_s = 1 + 6 + 15 + 20 + 15 + 6 + 1 \quad (\text{for } Y = 0, 1, 2, 3, 4, 5, 6, \text{ respectively})$$

$$C_s = 64 = 2^6$$

This value of 64 combinations available with the sequence code is to be compared with the 20 combinations offered as a maximum by the selector code.

The usefulness of the combinations available when using a sequence code is limited—as was the case with selector codes—by the fact that only one combination of the sequence code in question can be punched in any individual card. The reasoning for this is the same as was described in detail in considering selector codes. Punching two entries in a single sequence field of a card will—in general at least—establish a condition in which sorting operations will produce an intolerable number of extra cards. Consequently, sequence codes are principally useful for coding mutually exclusive concepts, which are to be sorted into a given sequence.

As already noted, a sequence code is the best possible one for arranging a group of cards in a predetermined sequential order. A minimum number of passes of the needle—equal to the number of holes in the field—is required (Chapter 2). Since—for a given number of coding combinations—a smaller number of holes is required for the sequence code than for either direct or selector codes, the superiority of the sequence code in this respect is obvious. To compute the amount of hand work required for arranging in sequence let us assume a sequence-coded field of H holes, a file of T cards which can be sorted 250 at a time. Then the number of passes of the needle needed for sequence sorting will be $TH/250$.

The efficiency of mechanical sorting based on a sequence code will be discussed next. The situation may be illustrated by considering a field of three positions (holes) in which eight numbers or letters may be coded.

Number coded	Positions punched		
	1	2	3
A		None punched	
B	X		
C		X	
D	X	X	
E			X
F	X		X
G		X	X
H	X	X	X

Assume that we wish to locate all cards on which letter F is coded. We will sort in position (hole) 3, and also drop out cards on which letters E, G, and H are punched. When we resort in position 2 in the cards previously dropped out, we will find cards coded for letters E and F remaining on the needle, and by resorting these cards in position 1 the desired cards coded for F will drop out. Evidently, making this sort required directing a search to all three positions. In like manner a search of the file for any of the other seven letters would also require needling all three positions. Generalizing, it may be shown that searching for any one combination punched in a sequence-coded field of H holes will require that all H positions be needed.

In setting up a sequence code for numbers the geometric series 1, 2, 4, 8, 16 . . . (binary code), or some variation thereof, can be used. The alphabet (26 items) may be accommodated by a field of five holes (total capacity 32 items). Such a field can be used to sort a group of cards into alphabetical sequence by only five passes of the needle. A very useful alphabetical mnemonic code O-I-E-C-B has been described by Casey, Bailey and Cox⁴.

The widely used 7-4-2-1 system for representing any digit by not more than two punches might be regarded as a variation of the binary code mentioned above.

Combinations of Fields

Up to this point we have discussed basic facts concerning different types of coding. In order to avoid undue complexity during this analysis of fundamentals attention has been directed to only one coding field. In practice, of course, combinations of fields are used.

A simple example of a code based on a combination of fields is furnished by the use of two modified sequence fields of four holes each to indicate a two-digit number. The four holes are labeled 7, 4, 2, 1 and coded as previously described in Chapter 2, page 19. In this way any number from 0 to 99, inclusive, may be indicated by eight holes. The same range of numbers could be covered by a field of seven holes using binary coding; 128 combinations would then be furnished. However, due to inexperience in thinking in binary most people find it more convenient to use the two modified sequence fields rather than a single seven-hole binary field. Consequently, the latter type of coding is not used ordinarily even though it requires fewer holes and also offers the advantage of permitting the cards to be arranged in sequence with fewer needlings. As far as C_n (coding combinations per needle) is concerned, the seven-hole binary coding is also superior with a C_n value of 128/7 or 18.3, as compared with the two-field code with a C_n of 100/8 or 12.5.

This simple example has been discussed in detail as it furnishes a good illustration of the conflict which sometimes arises between theoretical mathematical efficiency and considerations of a more practical psychological nature. Since we are accustomed to think in terms of a decimal system of numbers, the binary coding method confronts us with difficulties which may not be outweighed by its advantages, as mentioned above. Consequently, to achieve maximum usefulness and efficiency in the coding of multidigit numbers it is usually advisable to avoid binary coding and to code each digit separately in a 7, 4, 2, 1 type of field or some variation thereof.

⁴ Casey, R. S., C. F. Bailey, and G. J. Cox, *J. Chem. Ed.*, **23**, 495-9 (1946).

It is instructive to consider how combinations of either sequence or selector fields (or both) may be used to punch a multiplicity of entries in a single card. It is obvious that each combination of holes in each field may be assigned meaning. Hence, with the punching positions on a card divided up into a multiplicity of sequence or selector fields, a multiplicity of entries may be coded on the card. However, with the types of coding under discussion it is not possible to punch more than one entry in a given field. The reason for this is the previously noted fact that punching more than one entry per field will result in an excessive number of extra cards. As a consequence of this restriction, coding schemes based on a plurality of sequence and selector fields can scarcely be expected to give satisfactory results unless the analysis of the material to be coded can be based on sets of mutually exclusive concepts, examples of which are (1) the date at which some event (e.g., birth or death) occurred and (2) the location of an immovable object (e.g., a mountain). Often it is very inconvenient, if indeed not impossible, to conduct the analysis of information in terms of sets of mutually exclusive concepts. For this reason coding schemes based on combinations of sequence and selector codes have only very limited usefulness for many information problems. Fortunately, other coding methods, not subject to the above-mentioned restrictions, are available, and it is to them that we now direct our attention.

Direct Code Combinations

To illustrate this method of coding, let us suppose a biologist wished to use direct coding to record the usual classification of animals in terms of phylum, class, order, family, genus, species, and variety. Let us assume that we can divide, on some arbitrary basis, each phylum into 26 classes, each class into 26 orders, each order into 26 families, etc. In this way we would need only seven direct index fields of 26 punching positions each to classify an animal. This number of punching positions can be obtained on an edge-punched McBee "Keysort" card. Furthermore, each animal would require only seven punches to identify it.

The mathematics of such a system is of interest in connection with subsequent discussion of superimposed coding. Consider a card which listed only one animal. If we place a needle at random in the first or phylum field, there is only one chance in 26 that the needle will enter the position punched for the animal. Of course, we are assuming in these calculations that the distribution of punches is random or nearly so. If we needle all seven fields, there is only one chance in 26^7 or one chance in about 8,031,810,000 that all seven needles will correspond to the seven punched positions and allow the card to drop. It is convenient to define the dropping fraction (F_d)

as the chance that a given sorting operation (or series thereof) will cause a card taken at random to drop. For the case of needling the seven fields described above F_d would be the reciprocal of 8,031,810,000. Obviously, for this case, F_d would be a very small fraction indeed.

Superimposed Coding

It was pointed out in the previous paragraph how seven fields of 26 holes each might be used to code the classification of an animal in terms of classes, orders, etc. It was discovered that one punch in each of the seven fields would serve to indicate the classification of the animal, and that the dropping fraction, F_d , for a sorting operation involving seven needles inserted at random in each of the seven fields, would be $1/8,031,810,000$.

The dropping fraction increases rapidly as the number of animals coded on each individual card is increased. Thus, at first approximation, the effect of coding a second animal on each card would increase the probability to 2 chances in 26 that any single card would be selected by one needle placed at random in any of the seven fields. If seven needles were placed at random in each of the seven fields, there would be 2^7 chances in 26^7 (i.e., 128 chances in 8,031,810,000) of the card being selected. Obviously, the F_d in the case of two animals is $128/8,031,810,000$. Even if there were ten animals coded per card (ten punches per field), the dropping fraction would still be—at first approximation—only $(10/26)^7$ or $1/803$.

As indicated, the calculations of the preceding paragraph are approximate only. To make them more precise it is necessary to take into account the fact that the decision to code a second or third animal may not involve additional actual punching in all the fields. This can be made clearer if regarded in terms of our example involving animals. Assuming that they are equally distributed among the arbitrarily established 26 classes, there is one chance in 26 that the coding for the first animal will require us to punch a given position (e.g., position 4 in the class field). There is an equal chance (i.e., one in 26) that a second animal will require the same position (e.g., hole 4 in the class field) to be punched also. Or, to take a slightly different point of view, there are 25 chances in 26 that coding a second animal will require another position (i.e., some position other than 4 to be punched). After coding the second animal, $1 + 25/26$ positions rather than 2 will be punched in the class field. The number of unpunched positions will therefore be $26 - (1 + 25/26)$ or $24 + 1/26$, which can be written as $26(25/26)^2$. By similar reasoning it can be shown that for three subjects the number of positions left unpunched in the class field (or any other field) will be $26(25/26)^3$.

In more general terms, if X is equal to the number of subjects and H to

the number of positions in the field, then the number of positions left unpunched is

$$H \left(\frac{H-1}{H} \right)^X \quad \text{or} \quad \frac{(H-1)^X}{H^{X-1}}$$

If G is the number of positions punched, then it follows that

$$G = H - H \left(\frac{H-1}{H} \right)^X \quad (\text{Equation IV})$$

If in a field of H positions G positions are punched, then a needle inserted at random in this field will have G/H chances of entering a punched position. Using Equation IV, it can be seen that the value of G/H is very nearly equal to X/H when X , the number of subjects (e.g., animals) coded in a given field, is small. But for larger values of X , the value of G/H becomes significantly smaller than X/H . Thus, if H has the value 26 and X the value of 10, $X/H = 10/26 = 0.385$ and $G/H = 0.320$.

It seems appropriate at this point to develop a more convenient equation for calculating G/H , which is needed to calculate F_d , the dropping fraction previously defined.

Transposing Equation IV we write

$$\begin{aligned} \left(\frac{H-1}{H} \right)^X &= \frac{H-G}{H} \\ X \log \left(\frac{H-1}{H} \right) &= \log \left(\frac{H-G}{H} \right) \\ X &= \frac{\log \left(\frac{H-G}{H} \right)}{\log \left(\frac{H-1}{H} \right)} \end{aligned}$$

Now let G be some fraction of H , or fH . Then

$$X = \frac{\log \left(\frac{H-fH}{H} \right)}{\log \left(\frac{H-1}{H} \right)} = \frac{\log (1-f)}{\log \left(\frac{H-1}{H} \right)} \quad (\text{Equation IVa})$$

Now if H is large, this equation approaches

$$X = \frac{-\log (1-f)}{\frac{1}{H} \log e} \quad \text{or} \quad \frac{X}{H} = \frac{-\log (1-f)}{\log e} = \frac{-\log (1-f)}{0.434} \quad (\text{Equation IVb})$$

Table 21-2 gives the values of X/H , and f , the fraction of the field actu-

TABLE 21-2. NUMERICAL VALUES FOR EQUATION IV-b

The values of X/H are given in the body of the table, and f , the fraction of the field actually punched out is shown at the side and top of the table.

f	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.000	.010	.020	.030	.041	.052	.062	.072	.083	.094
.1	.105	.117	.128	.139	.151	.163	.174	.186	.198	.211
.2	.223	.236	.248	.261	.274	.288	.301	.315	.329	.343
.3	.357	.371	.386	.401	.416	.431	.446	.462	.478	.494
.4	.511	.528	.545	.562	.580	.598	.616	.635	.654	.673
.5	.693	.713	.734	.755	.777	.798	.821	.844	.868	.892
.6	.916	.942	.968	.994	1.02	1.05	1.08	1.11	1.14	1.17
.7	1.20	1.24	1.27	1.31	1.35	1.39	1.43	1.47	1.51	1.56
.8	1.61	1.66	1.71	1.77	1.83	1.90	1.97	2.04	2.12	2.21
.9	2.30	2.41	2.53	2.66	2.81	3.00	3.22	3.51	3.91	4.61

ally punched out, is indicated at the side and top. To illustrate use of this table assume that 9 entries have been punched in an alphabetically direct coded field of 26 positions. Assuming random distribution, how many positions are punched out and what is the dropping fraction? Since 9 names are coded in the 26 positions, the value of X/H is $9/26$ or 0.346. Looking in the table we find that it corresponds to a value of f , which is approximately 0.29. Therefore, the average number of positions punched out will be 0.29×26 or $7\frac{1}{2}$. The dropping fraction will be

$$\frac{G}{H} = \frac{fH}{H} = f = 0.29$$

This means that a needle inserted at random into the field of 26 positions with 9 entries punched will have 0.29 chances of entering a punched position. In other words, of every 100 cards in the file 29 would respond to our sorting operation. It will be recalled that our previous approximate calculation would have predicted $9/26$ or 0.346 chances of the needle entering a punched position (i.e., 35 cards per hundred in the file).

The considerations that led us to estimate 29 cards dropped per hundred if our hypothetical random search is directed to one field would predict that the same chance, namely 0.29, would apply to the same sorting operation directed to a second similarly punched field with a resultant probability of $(0.29)^2$ or 0.084 (i.e., about 8 cards dropping per 100 sorted). In general, for a search extending to Y similarly punched fields the dropping fraction, F_d will be

$$F_d = (G/H)^Y \quad (\text{Equation V})$$

Returning now to a consideration of the single needle inserted at random in a field of 26 positions into which 9 subject entries have been made, it is important to realize that the 29 cards selected out of a hundred on carrying

out a sorting operation include both the cards carrying pertinent information and those which do not. In other words, the dropping fraction is the sum of cards wanted (C_w) and extra cards (C_E)

$$F_d = C_w + C_E \quad (\text{Equation VI})$$

The problem of calculating C_w and C_E can be approached as follows: In searching a file which contains cards in which two alphabetic fields are coded as previously described, let us assume that all possible combinations of letters of the two alphabets (from AA to AZ and ZA to ZZ) are made on the cards and that two entries taken at random are made per card. Let us also suppose that the code combination AB is of particular interest to us. The dropping fraction F_d will be, using the approximate formula, $(2/26)^2$ or $4/676$. But, $(1/26)^2$ of the total combinations will be our desired AB; since there are two entries per card, there will be $2(1/26)^2$ or $2/676$ cards coded for AB. In this case the value of C_w is $2/676$ and, in view of Equation VI, C_E is

$$4/676 = 2/676 + C_E \text{ or } C_E = 2/676.$$

In other words, the cards selected mechanically would consist equally of wanted cards and extra cards.

In the general case involving Y fields of H positions, each punched to accommodate X entries, we would have

$$C_w = X \left(\frac{1}{H} \right)^Y \quad \text{and} \quad C_E = \left(\frac{G}{H} \right)^Y - X \left(\frac{1}{H} \right)^Y \quad (\text{Equation VII})$$

As a check, it is worth noting that when X is equal to one (i.e., only one entry in a field) G is also equal to unity and the expression for C_E becomes zero. With only one entry per field extra cards are, of course, impossible.

In the discussion of the last few paragraphs we have been considering a search directed to only a single coded entry—AB. In the more general case, we will wish to select cards pertaining to more than one subject, such as cards simultaneously punched for AB, BC, DE. If we are searching for M subjects, then the dropping fraction becomes

$$F_d = (G/H)^{YM}$$

Equation VII for wanted cards will be

$$C_w = X \left(\frac{1}{H} \right)^{YM}$$

and for extra cards

$$C_E = (G/H)^{YM} - X \left(\frac{1}{H} \right)^{YM} \quad (\text{Equation VIIa})$$

Since in most coding situations G is considerably greater than unity and X is at most only slightly larger than G , the C_x is very nearly equal to $(G/H)^{YM}$. In other words under our assumptions, most of the cards mechanically selected will be extra cards. In a final section of this chapter some experimental results relative to this point will be discussed.

Best Possible Utilization of Punching Positions in Superimposed Coding

In superimposed coding of the type which we have been considering in the previous section of this chapter, it was shown in the last of a series of conclusions that reducing the dropping fraction to a minimum is desirable as this would also reduce extra cards to a minimum (cf., Equations VII and VIIa). One factor that strongly influences the dropping fraction is the mode of dividing up our total number of holes into fields.

In the first edition of this book (Chapter 6) a generalized word coding system was described. This system employs combinations of four letters to code words and concepts. Typical code designations are SUGR for sugar, ACAD for Allied Chemical and Dye, IONE for ion-exchange, etc. These four-letter code designations were punched in four alphabetic fields, each of which consisted of 26 punching positions. This system illustrates one way in which 104 positions may be divided into fields when a scheme is set up for coding and punching. The 104 punching positions could have been divided up into two fields of 52 positions each, four fields of 26 positions each, or eight fields of 13 positions each. It is instructive to observe how the dropping fraction changes when a given number of positions is divided up into fields in different ways.

For a concrete example, let us assume that in each case 10 positions are to be punched in each field. This would be approximately equivalent to punching 10 subjects in each card, provided the codes are set up so as to require each subject to punch one position in each field (i.e., one punch in each of two fields for the two-field arrangement of 52 positions each, one punch in each of the four fields for the four-field arrangement of 26 positions each, etc.). The reasoning used in computing the data given in Table 6-2 of Chapter 6 of the first edition when applied to the problem under consideration produces the results given in Table 21-3.

TABLE 21-3
Number of positions in field divided by number of positions punched

Number of fields (Y)	Positions per field (H)	Number of positions in field divided by number of positions punched (H/G)	Dropping fraction (F_d)
2	52	5.2	$(10/52)^2 = 0.037$
4	26	2.6	$(10/26)^4 = 0.022$
8	13	1.3	$(10/13)^8 = 0.123$

Evidently the lowest value of F_d was obtained when four fields of 26 holes each were used. Some other grouping of the positions into fields will become associated with the minimum value of F_d if we make a moderately large change in the number of positions punched. Thus, if 5 positions are punched per field, the minimum F_d will be associated with a grouping of the 104 positions into eight fields of 13 positions; while if 15 positions are punched, the minimum F_d is associated with the grouping which involves two fields of 52 positions each. It turns out that the minimum value of F_d is associated with that arrangement of positions into fields for which the value of H/G is numerically most nearly equal to 2.718 (the exponential e). This "rule of e " can be demonstrated in a more general fashion as follows:

If we have a card with a total of C positions divided up into Y fields, each containing H positions of which G are punched and if we now use X to denote H/G , then $GX = H$ and, furthermore, $GXY = C$ since, by definition, $HY = C$.

Let us assume further that we have a series of different punching schemes, in which Y and H are permitted to change but always in such a way as to satisfy the equation $YH = C$. Let us also suppose that, no matter what the punching scheme, the value of G (i.e., number of positions punched in each field) remains constant. Then from the equation $GXY = C$, $Y = C/GX$ or, since both C and G are constant, we can substitute a constant K for C/G and obtain $Y = K/X$.

We next observe that if we place a needle in any one of the H positions of a given field there are G chances of the needle entering a punched position. Hence the ratio, R , of the number of cards needed in a single field to the number of cards dropping will be equal to H/G , which is equal to GX/G or X . For Y fields this ratio is X^Y . Since $Y = K/X$,

$$R = X^{K/X}$$

We next observe that R is the reciprocal of the dropping ratio F_d and, as already noted, we wish to discover what value of X will make R a maximum and F_d a minimum. Introducing logarithms we have

$$\log_e R = \log_e X^{K/X} = K/X \log_e X = K \frac{\log_e X}{X}$$

Applying the well-known formula:

$$d \frac{u}{v} = \frac{v du - u dv}{v^2}$$

$$d \log_e R = K \left(\frac{X X^{-1} dX - \log_e X dX}{X^2} \right) \quad \frac{d \log_e R}{dX} = K \left(\frac{1 - \log_e X}{X^2} \right)$$

If R passes through a maximum, its logarithm will pass through a maximum at the same point; therefore, we next equate the right-hand side of

the equation to zero to determine that value of X corresponding to a maximum value of R .

$$K \left(\frac{1 - \log_e X}{X^2} \right) = 0$$

This equation is true when $\log_e X = 1$ or when X is equal to e .

The rule of “ e ” may be stated in a slightly different form, as follows: The most efficient use of a limited number of holes is attained when the punching scheme is set up in such a way that, on an average, 37 per cent of the positions of each field is punched out. Numerically, 37 per cent is equal to $1/e$.

We have already pointed out earlier in this chapter that, assuming random distribution of punching, overlapping of punching can be expected to occur. As a result, the number of positions actually punched in a given field will be somewhat less than the number of symbols (e.g., letters) indicated by punching. Because of this overlapping effect, the optimum 37 per cent punching is attained when the number of symbols punched is equal to 46 per cent of the number of positions in the field. In practical terms this means that an alphabetical field of 26 punching positions attains our optimum 37 per cent punching when 12 letters have been coded in by punching, with random distribution being assumed as always. This same point will now be considered from a slightly different point of view.

First we recall Equation V, which permits us to calculate F_d for a search directed to one subject denoted by one punch each in a series of fields Y in number, with each field containing, on an average, a total of G punched positions. Then:

$$F_d = \left(\frac{G}{H} \right)^Y$$

Equation IV furnishes us with a relationship between H , the total number of punching positions in each of the Y fields; X , the number of subjects coded by one punched position in each field; and G , the number of positions punched on an average in each field.

$$G = H - H \left(\frac{H-1}{H} \right)^X$$

Or, dividing by H ,

$$\frac{G}{H} = 1 - \left(\frac{H-1}{H} \right)^X$$

Substituting in Equation V

$$F_d = \left[1 - \left(\frac{H-1}{H} \right)^X \right]^Y \quad (\text{Equation VIII})$$

TABLE 21-4. VALUE OF Y FOR A 144 HOLE SYSTEM WITH DIFFERENT VALUES OF X AND X/H

Subjects	Y					
30	1	2	2	3	3	4
20	2	3	3	4	5	6
19	2	3	3	4	5	6
18	2	3	4	5	6	6
17	2	3	4	5	6	7
16	2	3	4	5	6	7
15	2	4	4	5	7	8
14	3	4	5	6	7	8
13	3	4	5	6	8	9
12	3	4	6	7	8	10
11	3	5	6	7	9	10
10	4	5	7	8	10	12
9	4	6	7	9	11	13
8	4	7	8	10	12	14
7	5	8	9	12	14	16
6	6	9	11	14	17	19
5	7	11	13	16	20	23
4	9	13	17	21	25	29
3	12	18	22	27	33	38
2	18	26	33	41	50	58
1	36	53	66	82	100	115
Value of X/H	.25	.37	.46	.57	.69	.80

The form of this equation is such as to render computation of numerical values difficult. To save the reader's time, tabulations of values of F_d are given in Table 21-4 for various coding schemes based on a total of 144 punching positions. The values of F_d for the various coding schemes are given in terms of X (the number of entries coded) and X/H (the ratio of X , the number of entries coded, to H , the total number of positions in each field). Obviously, Y (the number of fields) is equal in each case to $144/H$ as we are basing our calculations on punching schemes involving a total of 144 holes. The number of fields (Y) for each of the punching schemes is given in Table 21-4. The values of Y are computed from the given values of X , the X/H ratio, and the relationship $Y = 144/H$.

From Table 21-4 it is clear that for any given value of X (the number of entries punched on the card) the dropping fraction (F_d) minimum shifts toward larger values of X/H as the value of X itself increases. These minimum values are italicized in Table 21-5, from which it is clear that, when X is unity and there is no chance of overlapping, the value of X/H is equal to $1/e$ or 0.37. As the numerical value of X is increased, the optimum ratio (shown by italicizing) passes through the value 0.46 and then very slowly

approaches the asymptotic limit of 0.69*. This shift in the minimum is caused by the overlapping effect of multiple punching, particularly when it occurs in a field consisting of a small number of punching positions.

It should be noted that Tables 21-4 and 21-5 refer to punching schemes of widely varying practicality. One way in which a punching scheme may become impractical is by requiring an excessive number of needles to search for a single subject entry. Since the number of needles for making a single search for a single subject is equal to the value of Y given in Table 21-4, it is clear that searching schemes falling in the lower right-hand corner of that table are scarcely practical. Furthermore, as shown in Table 21-5, the punching schemes set up for the case that X has a value of 30 have rather high dropping fractions, the most favorable value being 1.02×10^{-1} ; 10.2 per cent of the cards would drop on searching for any one subject. In setting up a coding scheme involving 144 punching positions we should not anticipate coding more than 20 entries per card, and we should divide up the 144 positions into five or six fields. In this way, the dropping fraction and hence the number of unwanted, extra cards is minimized. If we anticipate punching a smaller number of entries (on the average of 10 to 15 per card), then the dropping fraction will be appreciably improved by using a larger number of fields (eight or nine). The number of needles required to search for a given entry would, of course, increase as one is required for each field. This increase would tend to make the mechanical searching operations somewhat less convenient. In actually setting up a file a balance must be struck between these various factors. By multiplying the values shown in Table 21-5 by the corresponding number of needles shown in Table 21-4, one can obtain a new table giving the mechanical efficiencies, or the dropping fractions per needle. Such a table has shown that, on the average, the best results are obtained when the value of X/H is equal to 0.46.

In the preceding paragraphs and accompanying tables we applied Equation VIII to the analysis of the case that 144 punching positions are to form the basis of our system. Because of its general nature, Equation VIII can also be used to calculate dropping fractions for punching schemes based on a number of punching positions larger or smaller than 144. It turns out, as might be expected, that for a given dropping fraction the number of subjects that can be punched on the card increases and decreases in roughly direct proportion to the total number of punching positions involved. If, for example, we increase the total number of holes by a factor of five from 144 to 720, then the number of subjects that can be coded with a dropping fraction of less than 5 per cent increases from about 20 to about 100.

* This ratio of 0.69 corresponds to a value of $1/2$ for f in Equation IVb. From Equation II it was seen that when $1/2$ of the holes were punched out, or $f = 1/2$, the number of combinations was a maximum.

TABLE 21-5

X	1.64×10^{-1}	1.26×10^{-1}	1.12×10^{-1}	1.04×10^{-1}	1.02×10^{-1}	1.05×10^{-1}
30						
20	6.68×10^{-2}	4.50×10^{-2}	3.79×10^{-2}	3.41×10^{-2}	3.34×10^{-2}	3.47×10^{-2}
19	5.80×10^{-2}	3.83×10^{-2}	3.20×10^{-2}	2.86×10^{-2}	2.80×10^{-2}	2.92×10^{-2}
18	4.95×10^{-2}	3.20×10^{-2}	2.65×10^{-2}	2.36×10^{-2}	2.31×10^{-2}	2.41×10^{-2}
17	4.15×10^{-2}	2.62×10^{-2}	2.14×10^{-2}	1.89×10^{-2}	1.86×10^{-2}	1.95×10^{-2}
16	3.41×10^{-2}	2.09×10^{-2}	1.69×10^{-2}	1.49×10^{-2}	1.46×10^{-2}	1.54×10^{-2}
15	2.72×10^{-2}	1.62×10^{-2}	1.29×10^{-2}	1.13×10^{-2}	1.11×10^{-2}	1.17×10^{-2}
14	2.11×10^{-2}	1.21×10^{-2}	9.50×10^{-3}	8.26×10^{-3}	8.09×10^{-3}	8.63×10^{-3}
13	1.57×10^{-2}	8.62×10^{-3}	6.68×10^{-3}	5.75×10^{-3}	5.64×10^{-3}	6.06×10^{-3}
12	1.11×10^{-2}	5.83×10^{-3}	4.43×10^{-3}	3.78×10^{-3}	3.71×10^{-3}	4.03×10^{-3}
11	7.41×10^{-3}	3.67×10^{-3}	2.75×10^{-3}	2.30×10^{-3}	2.27×10^{-3}	2.49×10^{-3}
10	4.56×10^{-3}	2.11×10^{-3}	1.53×10^{-3}	1.27×10^{-3}	1.26×10^{-3}	1.41×10^{-3}
9	2.52×10^{-3}	1.08×10^{-3}	7.55×10^{-4}	6.21×10^{-4}	6.22×10^{-4}	7.07×10^{-4}
8	1.20×10^{-3}	4.64×10^{-4}	3.14×10^{-4}	2.55×10^{-4}	2.58×10^{-4}	3.02×10^{-4}
7	4.63×10^{-4}	1.58×10^{-4}	1.02×10^{-4}	8.18×10^{-5}	8.46×10^{-5}	1.03×10^{-4}
6	1.31×10^{-4}	3.81×10^{-5}	2.33×10^{-5}	1.84×10^{-5}	1.96×10^{-5}	2.54×10^{-5}
5	2.25×10^{-5}	5.28×10^{-6}	3.00×10^{-6}	2.55×10^{-6}	2.67×10^{-6}	3.80×10^{-6}
4	1.64×10^{-6}	2.81×10^{-7}	1.47×10^{-7}	1.17×10^{-7}	1.54×10^{-7}	2.56×10^{-7}
3	2.16×10^{-8}	2.35×10^{-9}	1.12×10^{-9}	9.82×10^{-10}	1.73×10^{-9}	4.33×10^{-9}
2	4.55×10^{-13}	2.44×10^{-13}	1.19×10^{-13}	1.74×10^{-13}	8.54×10^{-13}	6.85×10^{-13}
1	2.12×10^{-23}	9.85×10^{-24}	4.39×10^{-23}	9.16×10^{-21}	1.29×10^{-16}	6.85×10^{-13}
Subjects	0.250000	0.367879	0.458676	0.570000	0.693147	0.800000
			Value of X/H			
		144 Holes	Dropping Fractions or F_d			

Summarizing, we may say that for practical punching schemes involving a maximum of entries punched on the card the optimum conditions are attained when the number of instructions to punch in a given field amounts to approximately 46 per cent of the available punching positions. Because of overlapping, such a proportion of punching instructions will result in approximately 37 per cent ($1/e$) of the positions being actually punched.

Other Variations of Superimposed Coding

In addition to the multiple direct code combination described above, there is a somewhat similar method of coding which has been used by the McBee Co. for various installations of "Keysort" cards. This coding procedure arbitrarily assigns to each index entry, such as "nonsugar", some combination of punching positions. Each of these combinations consists of a small fixed number of holes (e.g., four) selected at random from the totality of punching positions being used. By avoiding restrictions imposed by fields on the randomness of the punching pattern, single-field superimposed coding (as we shall term this coding method) attains somewhat more favorable dropping fractions when other conditions (viz., number of needles, Y , number of subjects, X , and total number of holes used) remain the same. To calculate precisely the numerical values of the dropping fractions, the following equation* should be used,

$$F_d = \frac{\frac{G!}{Y!(G-Y)!}}{\frac{H!}{Y!(H-Y)!}} \quad (\text{Equation IX})$$

where, as before

$$G = H - H \left(\frac{H-1}{H} \right)^X$$

To explain the meaning of Equation IX a specific example may prove helpful. Assume that a 144-hole field is to be used in this manner and that there are an average of sixteen subjects to code and a 4-needle punching scheme to be used. Then $4 \times 16 =$ a total of 64 punches would be required, some of which would overlap. To obtain the number of actual punches a modified form of Equation IV should be used. The modification is necessary because in this case the 4 needles are all punched in the same field, instead of only one needle per field. The altered formula would be

* Mooers, in discussing this type of coding—which he terms "Zatocoding"—has used the less precise equation, $F_d = (G/H)^Y$. (Paper by C. N. Mooers presented before the Division of Chemical Education at the 112th meeting of the American Chemical Society, New York, N. Y., Sept., 1947.)

$$G = H - H \left(\frac{H - Y}{H} \right)^x$$

Sine $Y = 4$ in this case, the equation would give

$$G = 144 - 144 \left(\frac{144 - 4}{144} \right)^{16} = 52$$

Now in order for the card to drop all four needles must be inserted in this particular group of 52 punched holes. By the use of Equation I we know that there are

$$\frac{H!}{Y!(H - Y)!} = \frac{52!}{4!48!} = \text{ways this can happen.}$$

But Equation I also tells us that the total number of ways the 4 needles can be inserted in the entire 144-hole field is

$$\frac{144!}{4!140!}$$

Therefore, the dropping fraction will be the number of ways the needles can be inserted in the 52 punched-out holes divided by the number of ways the needles can be inserted in the entire field of 144 holes, or

$$F_d = \frac{\frac{52!}{4!48!}}{\frac{144!}{4!140!}} = 1.56 \times 10^{-2}$$

Table 21-6 gives numerical values of F_d calculated with the aid of this equation for the case that $H = 144$. Y , the number of needles, is varied in accord with the equation $Y = 144/H$, as given in Table 21-4. Very similar trends will be observed with respect to numerical values of F_d in Table 21-6, as already noted in Table 21-5. However, it should be noted that the minimum optimum value of X/H is 0.500000 or $(1/2)$ instead of 0.367-879 $(1/e)$ in Table 21-5. Both systems of punching have the value of X/H approaching a maximum optimum of 0.693147 as the number of subjects increases.

A further variation of single-field superimposed coding was developed by Isbell⁵. His system is characterized by use of a 3-needle code for a main classification heading together with a 1- or 2-needle code for each minor classification. This arrangement of the code complicates calculation of the

⁵ Isbell, A. F., "A Practical Application of a Punched-Card System Utilizing the Superposition of Codes", paper presented before the Division of Chemical Literature of the American Chemical Society at the 114th national meeting in Portland, Oregon, Sept., 1948.

dropping fraction. For mathematical details concerning Isbell's very excellent coding system the reader is referred to his original paper.

Another variation of superimposed coding was developed by Guy and Geisler⁶. Their system was a very simple and effective author index consisting of two alphabetical fields of 26 letters each. The first letter was recorded by a deep punch in the appropriate section and the second by a shallow punch. A "deep" punch is actually a combination of a shallow and an intermediate punch. Therefore, the first field was filled up with coding twice as fast as usual since it included the deep punching of the first letter plus the shallow punching of the second. The second field was punched out at the normal rate as it was affected only by the deep punch of the first letter of the author's name. Except for this excessive punching of the first field, the Guy and Geisler superimposed coding had all the characteristics of an ordinary two-field system of 26 letters each*.

Comparison of Actual and Theoretical Dropping Fractions

In our previous discussions concerning dropping fractions we have been mainly concerned with theoretical rather than actual values although it was stated that the actual mathematical behavior might be quite different from the theoretical. It is very difficult to make definite statements concerning actual behavior because each file will be different. However, some general trends can be pointed out.

According to Keckley¹ it has been found that in general record keeping with any extensive system of classifications there is a central tendency for 90 per cent of the activity to be concentrated within 25 per cent of the classifications. This would mean that if a simple code of 100 ideas is used and if we needle for one of the 25 per cent group, our dropping fraction will not be 1/100 but will average $1/25 \times 90$ per cent, or 1/28. On the other hand, if we needle for one of the 75 per cent group, our dropping fraction would average $1/75 \times 10$ per cent, or 1/750.

It was to avoid this unevenness of efficiency that led Cox, Casey, and Bailey² to divide an alphabetical arrangement of initial letters of authors' names into 100 even divisions according to frequency of occurrence, and then number these divisions consecutively.

For the superimposed type of coding there have been at least two independent studies on the relationship between calculated and actual dropping fractions. One was an unpublished study by the author and the other the excellent paper by A. F. Isbell⁵; both gave similar general conclusions.

Isbell was using a modification of the single-field superimposed coding

⁶ Guy, A. G., and Geisler, A. G., *Metal Progress*, 52, 993-1000 (Dec. 1947).

* Guy and Geisler used this combination deep- and shallow-punching scheme so that the cards could fall completely free of the needles. The use of an "intermediate" punch alone is generally not as satisfactory as a shallow punch.

method. The modification consisted in dividing his random code pattern into subclassifications, and this in turn modified the dropping fractions. For example, if a 5-needle code is used for a 58-hole single field and codes 5 subjects, the theoretical sum of the code punches is 25 and the corresponding G will be about 20. But if a 3-needle code is used for the main subjects and a 2-needle code for the subclassifications, then the G will be reduced because many of the subclassifications will belong to the same main classification. As an example, a card might include five subjects, all of which deal with the same general classification. Then the sum of the code punches would be $3 + (2 + 2 + 2 + 2 + 2)$, or only 13, and the corresponding G would be less than this. A much better dropping fraction would result. On the other hand, if one tried to sort for a minor subject or combination of minor subjects by themselves, then the system would be less efficient than if the regular single-field type of code were used.

For the above reason, Isbell found his actual G to be less than the theoretical by about 11 per cent. When the author experimented with his filing system, he found his actual G to be about 5 per cent lower than the theoretical value. The reason for the latter was that an alphabetical code was being used, and, as previously mentioned, the alphabet is not a random system.

Both Isbell and the author found that the total dropping fraction is about 100 times the theoretical value. This is due to the fact that in actual practice one sorts normally for only those code combinations that are represented by at least one card in the file, and the very large number of undesignated codes are never sorted for. Since the total dropping fraction includes both desired and extra cards, the 100 to 1 discrepancy is mainly due to the desired cards and is an advantage, rather than otherwise. Doubtless for larger and less specialized files the difference between the actual and theoretical total dropping fractions would be much less.

On the other hand, both workers found that the actual *extra card* dropping fractions tended to be fairly near the theoretical values. For example, the author, in a study of the coding of scientific authors, found that the 26 letters of the actual alphabet gave an extra card-dropping fraction equivalent to a theoretical random system of about 19 letters.

It should be pointed out here that it is possible to expand the alphabet into a statistically random "alphabet" of 30, 100 or even 1000 positions per field. The clearest description of the general idea is given in a discussion of triangular methods of alphabetical coding by Cox, Casey and Bailey.² In this same paper they also use an "alphabet" shorter than 26 letters by combining some of the less frequently used letters. Thus the 26 letter alphabet, which admittedly is not random, may either be expanded or contracted, and in either case the resulting modified "alphabet" approaches the theoretical random type. As indicated in the paragraph above, this modifica-

tion of the alphabet to give a more random distribution has not been found necessary in word coding with hand sorted cards. However, it might be useful when used on a larger scale, as for example, in the Rapid Selector. For a more complete discussion of this, the reader is referred to the author's "Multiple Word Coding *vs.* Random Coding for the Rapid Selector. A Reply to Calvin N. Mooers." *Am. Documentation*, 3, No. 4, 223-225 (1952).

Some Comparisons

By this time the reader may have asked himself, "Which is the best method of coding?" This question cannot be answered unequivocally because each of the four main methods of coding has its own peculiar advantages and disadvantages. As Isbell has so aptly stated, "... it soon becomes obvious to anyone working in this field that there are a great many modifications of the punched-card systems that have been described, and the factors controlling the effectiveness of any one system are many and intricately interwoven". The reader may, therefore, design from the foregoing a system which best suits his needs.

As pointed out in Chapter 18, punched-card codes, to have maximum effectiveness, must be built up from those terms and concepts most useful as a basis for performing sorting operations when using the punched-card file as a source of information. The nature of the terms and concepts which make up the code is an important factor influencing the punching scheme to be developed. Almost equally important in this regard is the total number of terms and concepts involved in the code. If the total number of terms and concepts forming the code is smaller than the number of punching positions on the card and if much sequence sorting is not necessary, then a direct code is probably the simplest and best.

If the total number of terms and concepts exceeds the number of available punching positions, then it is necessary to use a punching scheme which attributes meaning to combinations of positions. In this case, the nature of the terms and concepts will usually be a controlling factor with regard to the punching scheme to be selected. Terms and concepts of a mutually exclusive nature can be grouped together and assigned to fields set up for selector or sequence punching. However, as already noted, such punching schemes are restricted to only one entry per field. In order to provide for a code involving a large number of terms and concepts which are not mutually exclusive in nature, recourse must be had to superimposed coding. In arriving at a decision concerning which type of superimposed coding to use the following considerations may prove helpful.

We have seen that single-field superimposed coding gives slightly better dropping fractions than the corresponding multiple-field method. Figure 21-2 shows this graphically. In this figure the logarithm of the dropping fraction is plotted against the punching ratio. It is to be noted that the

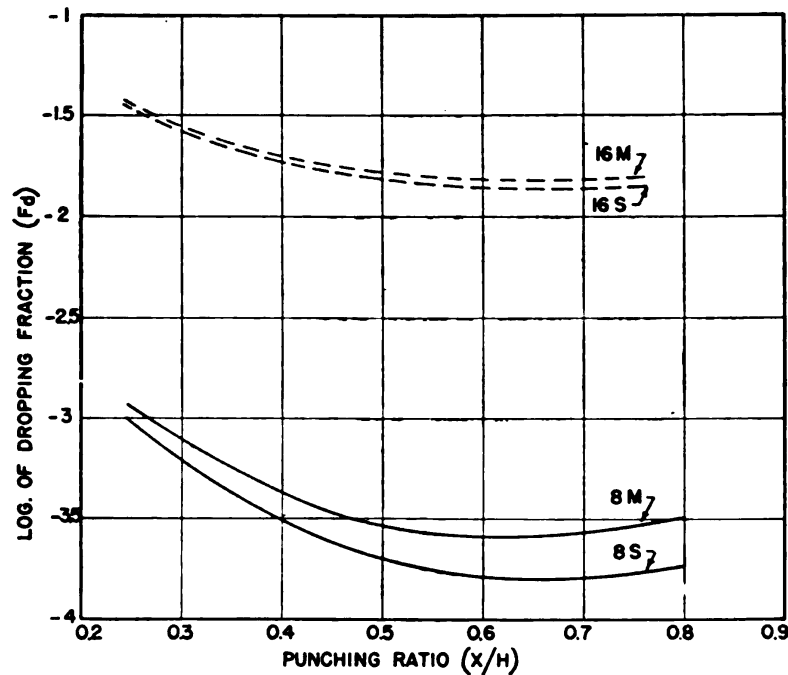


FIGURE 21-2. Effect of punching ratio change on dropping fraction.

dropping fractions for the two systems approach each other as the number of subjects coded increases. For example, the dotted line, marked "16M", which refers to the dropping fraction for the multiple-field method using 16 subjects, is very close to the single-field set of dropping fractions for 16 subjects, marked "16S". For the unbroken lines, marked "8M" and "8S", which refer respectively to the multiple- and single-field methods with only 8 subjects coded, the difference is more marked. However, in this latter case both sets of values of F_d are so low that any difference is not really significant. As a specific example, it is hard to detect any practical difference between an F_d of $1\frac{1}{2} \times 10^{-4}$ and one of $2\frac{1}{2} \times 10^{-4}$.

This minor superiority of the single-field method with regard to the dropping fraction is, in actual use, more than balanced out by the fact that the multiple-field method permits the code designations to be built up on mnemonic principles. In practice, this means that a multiple-field system is much more convenient to use. The situation is closely analogous to the relationship between coding numbers by using a binary system or by coding each decimal digit separately. Considered solely from the standpoint of efficient use of the punching positions on the card the binary system is appreciably more efficient. However, from the standpoint of convenience in actual use the digital coding method is far superior.

Another advantage of the multiple-field method of superimposed coding is the fact that the fields may be numbered and significance may be attributed to the order in which the fields are used when punching a given coding

designation. Thus, to use a homely example, the multiple-field method can be used to distinguish different relationships between "man" and "dog" in the two sentences "Man bites dog" and "Dog bites man". Using a three-field arrangement, the first sentence might be coded as follows:

Field					
I	M		T		O
II	A	B	E		G
III	N	I	S	D	

While the other case in which the man is bitten might be coded as follows:

Field					
I	D		T		A
II	O	B	E		N
III	G	I	S	M	

"Dog bites man".

In coding, the subject of the sentence starts with Field I and proceeds through Fields II and III, while the object starts with Field III and proceeds through Fields I and II in that order. This illustrates two possible ways to code "man", "dog", or any other three-letter word or code. Obviously, such three-letter words or codes could be punched in many different ways (e.g., by starting with Field II and proceeding to Field I and later to Field III).

One practical application of these possibilities is in recording the Dyson code for organic structural formulas. The Dyson code itself can be considered as a type of sentence in which not only the words (symbols) but also their relative positions are important. In this respect it is like our Arabic numeral system, where the number 123 has quite a different meaning from the number 321. Other applications of these possibilities remain to be worked out.

Conclusion

Punching schemes developed to date are characterized by a number of different features. Their most advantageous exploitation when setting up a file for a specific purpose is impossible without careful thought, and may well require a certain amount of preliminary experimentation.

It is believed that the mathematical principles developed in this chapter are valid for many types of mechanical searching devices,^{7,8} including complex electronic equipment (e.g., "electronic brains"). The possibilities inherent in the new approach to the analysis of information have hardly been scratched. The "Memex" of Vannevar Bush⁹ may be closer than is generally realized.

⁷ Wise, C. S., and Perry, J. W., *Am. Documentation*, 1, No. 2, 76-83 (1950).

⁸ Wise, C. S., *Am. Documentation*, 3, No. 4, 223-225 (1952).

⁹ Bush, V. *Atlantic Monthly*, 176, 101-8 (July, 1945).

Chapter 22

COMPREHENSIVE CODING SCHEMES FOR CHEMICAL COMPOUNDS

D. E. H. FREAR

The Pennsylvania State University, State College, Pa.

Introduction

In the past, chemical compounds have been indexed and classified according to the obvious, time-tried bibliographic methods of (a) alphabetical arrangement of names, (b) empirical formulas, or (c) arbitrary type classification. At one time only a few hundred compounds were known, and any one of these was satisfactory. With the tremendous growth of synthetic organic chemistry within recent years—a conservative estimate of known organic compounds is at least 500,000—none of these methods proved workable, and, as a result, much valuable scientific information is either inaccessible or totally lost for all practical purposes.

The alphabetical arrangement of names is probably the least desirable method of indexing and classifying chemical compounds since, in many cases, several names may be applied to a compound with equal correctness. Certain rules of nomenclature have been laid down, and in the United States those used by the staff of *Chemical Abstracts* are generally accepted as standard. While these are usually clear and relatively easy to follow, in the more complex compounds there have been instances where confusion has existed. Chemists in other countries have rules of their own for naming compounds, and this naturally adds to the confusion of the worker who is, for example, searching for all available information on a given compound.

Empirical formulas are useful as a basis for indexing, but are subject to certain limitations. Disregarding the possibility of error in totaling the number of atoms in a compound (which may be quite likely in large, complex compounds), the main objection to empirical formulas is their lack of information. They serve to arrange compounds in an ordered sequence, but give no information on the relationships between compounds or on what compounds may be of common derivation. The empirical formulas for relatively simple compounds are not specific; many different compounds having the same empirical formula may occur.

Classification of chemical compounds by types is useful for certain purposes. Where relatively small numbers of compounds are involved or where a group of compounds has a common denominator, it is feasible to set up

type classes which are useful for indexing purposes. Any attempt to set up a type classification system for all chemical compounds is extremely difficult. For example, let us say that we are going to divide all organic compounds into broad classes—the halogen derivatives, the amines, the heterocyclic compounds, etc. Immediately a question is raised as to the disposition of a compound in which two or more of these groups are present. The solution involves setting up priority lists so that halogens come before amines, for example, and all compounds having both halogens and amines, thereby, are classified as halogen derivatives. The consequence is that each of the early classes will contain representatives of all of the later groups, which may be lost for all practical purposes unless an elaborate cross-indexing system is worked out.

Classification and indexing of chemical compounds is not always the sole ultimate objective of the chemist. With the mass of scientific data available it is often desirable to investigate all compounds having common groups, or those derived from one parent compound. Frequently, these structural features must be studied in relation to physical, chemical, or biological properties. For this purpose, the name and the empirical formula are useless, and another system must be devised which will make correlation studies possible. Although the use of a code for correlation studies is quite a different matter from classification, the two are closely related and many chemical code-makers have attempted to combine the two objectives with varying degrees of success.

Some workers have suggested that chemical codes may also be used as a means of communication between chemists, replacing formulas and names of oral or written exchanges of information.

These three divergent points of view have led to some confusion in the field of chemical coding. It may well be that no one system will ever accomplish all objectives: indeed, there is a general opinion in some quarters that there may be need for not one, but several types of chemical codes, depending on the specific application required.

The large number of chemical compounds now known makes it imperative that any chemical coding system be adaptable to mechanical manipulation. Punched cards offer a ready means of sorting, classifying and correlating large numbers of individual pieces of data, and are ideally suited for tasks of this sort. Electronic searching and calculating machines such as UNIVAC or IBM 700 type have been developed in recent years for the purpose of computing and correlating masses of data. These complex machines, when supplied with suitable information, can compare structural units and determine relationships between such units in molecules. With such machines a relatively simple atom-by-atom delineation of the compound may be sufficient, thus eliminating the need for designating structural or functional groups.

The various chemical coding schemes which have been proposed fall into three broad categories. In the first class (a) are those which are unique (only one cipher for a given compound) but ambiguous (the same cipher may apply to more than one compound). Examples of this type are the Frear, CBCC, and Wiselogle codes. The second class (b) includes those in which the ciphers are unique and unambiguous, such as the Dyson, Wiswesser, and Gruber codes; and the third class (c) are those in which the ciphers are not unique but are unambiguous, such as the systems of Opler and Norton, the U. S. Patent Office and others. Codes in class (a) have demonstrated their usefulness in correlation studies involving structural features, especially when used with punched cards, and in a limited way in indexing and classifying. For example, the CBCC code has been applied to approximately 70,000 compounds to relate structure with biological activity; the Frear code has been used to relate structure to insecticidal efficiency in some 10,000 compounds. The completely specific codes in class (b) promise to be useful for structure searching, indexing and classifying; the codes in class (c), requiring fewer rules but more complicated machinery, are still in the development stages. It appears that they can be useful for structure searching, especially where large numbers of compounds are involved.

The present trend seems to be toward the development of specific coding systems for particular applications. It is probable that the choice of the code to be used will depend on whether the purpose is indexing, classifying, correlating or structure searching, and the availability of mechanical equipment, such as punched card sorters, digital computers, etc. For example, as more versatile equipment becomes available, coding systems, such as that of Opler and Norton, are being developed to take advantage of the machines. Thus we may expect new, and perhaps better coding systems for chemical compounds in the future, although certain of the existing systems are quite satisfactory for the purposes for which they were designed.

A Commission of the International Union of Pure and Applied Chemistry (IUPAC) has been studying the linear notation problem since 1946, and has done much to encourage work along these lines. In 1949 this Commission set up the following desiderata for evaluating notation systems:

1. Simplicity of usage.
2. Ease of printing and typewriting.
3. Conciseness.
4. Recognizability.
5. Ability to generate a unique organic chemical nomenclature.
6. Compatibility with accepted practices of inorganic chemical notation.
7. Uniqueness.
8. Generation of an unambiguous and useful enumeration pattern.
9. Ease of manipulation by machine methods, e.g., punched cards.

10. Exhibition of associations (descriptiveness).

11. Ability to deal with partial indeterminants.

At its 1949 meeting the Commission decided to invite the inventors of notation systems to submit them for consideration, and nine were so submitted. A list of over 700 compounds was submitted to the inventors, who prepared notations for each compound on the list.¹ The IUPAC Commission, after considering this survey, decided that the Dyson system came nearest to meeting the desiderata, and this system was adopted as a provisional international standard. The suggestion was made at that time that it might be possible to incorporate some of the desirable features of other systems into the Dyson notation.

In 1952 a committee of the National Research Council distributed a representative sample of structural formulas to a group of over 100 volunteer chemists for enciphering and deciphering. These volunteers compared the Dyson, Gruber, Silk, and Wiswesser notations. Their findings indicated that the number of disagreements in the enciphering process varied from 50 to 70 per cent; disagreements in the deciphering operation varied from 18 to 26 per cent. The average time required to encipher 1000 compounds was between 115 to 192 hours; to decipher the same number required from 50 to 67 hours². The volunteers reported that most of the notation systems lacked completely specific directions. In the light of these results several of the authors have modified their codes, and at least one new system has been proposed³ which is reported to include the best features of the previously existing codes.

In the following section of this chapter a brief description will be given of the existing systems for chemical coding. The space available precludes an exhaustive discussion of any particular code, but in most instances books, pamphlets or technical articles are cited in the bibliography which will give the interested reader further information.

The Frear Code^{4,5}

This code for assigning numerical designations to chemical structures was devised by Frear, Seiferle, and King⁶ to enable them to study the

¹ Berry, Madeline M., and Perry, J. W., Notational Systems for Structural Formulas. *Chem. Eng. News*, 30, 407-410 (1952).

² Staff Report, *Chem. Eng. News*, 33, 2838-43 (1955).

³ Crane, E. M., and Berry, Madeline M., The composite volunteer notation system for molecular structural formulas. Paper presented before the Division of Chemical Literature, American Chemical Society Meeting, Cincinnati, Ohio, April 2, 1955.

⁴ Frear, D. E. H., *Chem. Eng. News*, 23, 2077 (1945).

⁵ Frear, D. E. H., "A Catalogue of Insecticides and Fungicides," Waltham, Mass., The Chronica Botanica Co., 1948.

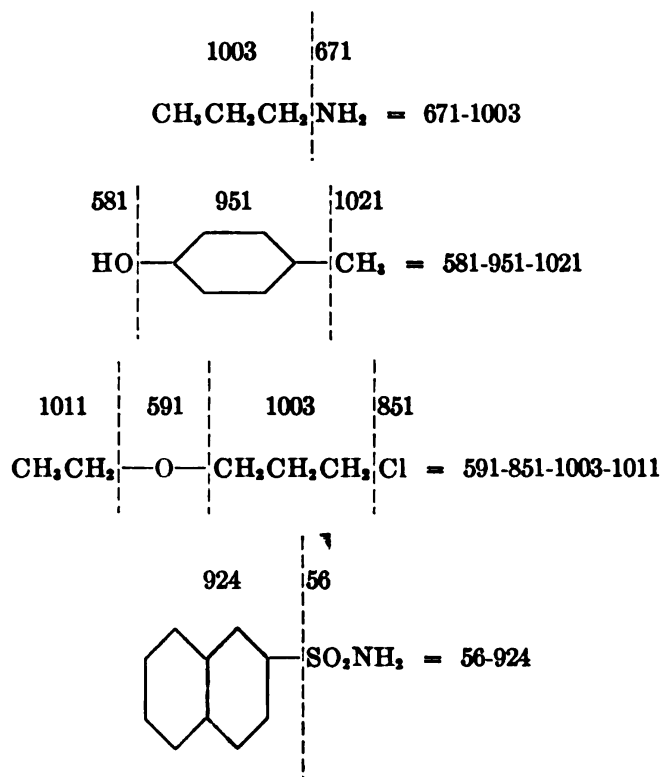
⁶ Frear, D. E. H., Seiferle, E. J., and King, H. L., *Science*, 104, 177-178 (1946).

correlation between biological activity and chemical constitution in a group of more than 6,000 compounds of a widely varied nature⁷. A complete description of this system has been published⁵, and only a brief resumé will be presented here.

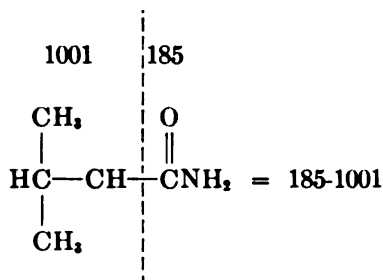
In this system of coding the emphasis is placed on the structural units of the chemical compounds. Approximately 400 constituent groups were selected to cover organic structures, and roughly half as many for inorganic groups. To each of these was assigned a number, the groups being arranged in decreasing order of complexity.

To assign a code number to a particular compound the list of constituent groups is read downward, starting with the most complex; as a group present in the compound is encountered in the list, the number is noted, and the process continued until all constituent groups have been accounted for. For example, in ethyl alcohol the first group encountered is the hydroxyl, which bears the group number 581. The remainder of the compound, the two-carbon chain, is farther down the list, and bears number 1011. Therefore, $\text{CH}_3\text{CH}_2\text{OH} = 581-1011$.

Similarly, the following compounds are assigned code numbers, as indicated: (Dotted lines separate the constituent groups).



⁷ Frear, D. E. H., and Seiferle, E. J., *J. Econ. Entomol.*, **40**, 736-741 (1947).



In the fourth example the sulfonamide group, $-\text{SO}_2\text{NH}_2$, is numbered 56; while $-\text{SO}_2-$ and $-\text{NH}_2$ individually bear higher numbers, 264 and 671, respectively, so that the more complex group is coded as a unit and not split into smaller parts. The arrangement of groups in decreasing order of complexity is the basis of this system of coding, and for most compounds coding is a relatively simple process. The common inorganic anions and cations are coded in the same way so that a simple inorganic salt will have two group numbers.

This relatively simple code has several advantages. It has been used successfully to code over 10,000 compounds, and the procedure may be learned easily by anyone with a minimum of chemical knowledge. The code numbers may be used on either hand-sorted or machine-sorted punched cards, the former being preferred for small operations.

There are, however, certain disadvantages and shortcomings in the system. With a short list of constituent groups it is obviously not possible to make provision for all theoretically possible combinations. This is particularly true of inorganic compounds and out-of-the ordinary metallo-organic derivatives. In short, while this system proved adequate for the purpose for which it was designed and will accommodate 99 per cent of the compounds commonly encountered, it will not designate certain structures and combinations without extensive additions and alterations. The modifications discussed in the following section were designed to correct these shortcomings.

National Research Council Code^{8, 9, 10}

Faced with the problem of correlating chemical structure with biological activity for a large number of compounds, the Chemical Codification Subcommittee of the National Research Council began work on a code for chemical compounds in 1944. Because of the large number of chemicals

⁸ Bailar, J. C., Jr., Heumann, K. F., and Seiferle, E. J., *J. Chem. Ed.*, **25**, 142-144 (1948).

⁹ Chemical Codification Panel, National Research Council, "A Method of Coding Chemicals for Correlation and Classification," Washington, D. C., 1950.

¹⁰ Morgan, J. A., and Frear, D. E. H., *J. Chem. Ed.*, **24**, 58-61 (1947).

involved and the nature of the information desired, two fundamental requirements were set up: The code should be specifically designed for machine-sorted punched card use, and it should indicate constituent groups in the compounds for correlation purposes. After examining the systems then in existence, the subcommittee decided to adapt the code devised by Frear and associates (see preceding section) to meet the specific requirements of the problem at hand.

The basic principles of the Frear⁵ code were retained in the N.R.C. code. For greater flexibility, several major changes were made in the original system, and it is the opinion of the committee that the code in its final form will accommodate all compounds in a logical manner. As part of the broad project undertaken by the National Research Council, a related code has been developed for biological properties so that correlation studies may be made by simple operations of punched-card machines.

The principle of numerical designation of constituent chemical groups has been retained in the N.R.C. code. The code number for a given compound is made up of a series of four-digit group numbers, each representing a specific structure in the molecule. The first character in the group number designates the family to which the structure belongs, that is, whether it contains, for example, carbon, hydrogen, nitrogen, oxygen, and sulfur, or carbon, hydrogen, nitrogen, and oxygen, or carbon, hydrogen, and nitrogen in organic groups. The family designations are given in Table 22-1. The second and third digits in the group number identify the particular chemical structure (for example, a nitro or hydroxyl group), and the fourth digit denotes the number of times the group occurs in the compound or ion.

In Division A, or the "organic" section of the code, the families are listed in order of decreasing complexity with respect to the number of different elements in the group, proceeding from Family 0, which contains carbon, hydrogen, nitrogen, oxygen, sulfur, and halogen, to Family Ø*, which contains only carbon and hydrogen. Generally, two families are used for each combination of elements, one for noncyclic groups and the other for cyclic groups. The noncyclic family precedes the corresponding cyclic family for each combination of elements.

In the "organoheteroid" section, Division B, any element other than C, H, N, O, S, or X attached directly to carbon is coded in Family P. The specific element and its combining power are described by the second and third characters of the group number.

Division C, the "inorganic" part of the code⁸, contains families Q to V. These latter include carbonless (Q) ring structures, free elements, simple cations, and central elements of complex cations or neutral molecules (R),

* Ø is used to indicate the letter O, to avoid confusion with zero.

TABLE 22-1. LIST OF FAMILIES

<i>Division</i>	<i>Family</i>	<i>Composition of Family</i>
I	0—	(CH)NOSX
	1—	(CH)NOS—Noncyclic groups*
	2—	CNOS(Z) Rings
	3—	(CH)NOX
	4—	(CH)NSX
	5—	(CH)OSX
	6—	(CH)NO—Noncyclic groups*
	7—	CNO(Z) Rings
	8—	(CH)NS—Noncyclic groups*
	9—	CNS(Z) Rings
	A—	(CH)OS—Noncyclic groups*
	B—	COS(Z) Rings
	C—	(CH)NX
	D—	(CH)OX
	E—	(CH)SX
	F—	(CH)N—Noncyclic groups*
	G—	CN(Z) Rings
	H—	(CH)O—Noncyclic groups*
	I—	CO(Z) Rings
	J—	(CH)S—Noncyclic groups*
	K—	CS(Z) Rings
	L—	(CH)X
	M—	CZ Rings
	N—	C Rings
	Ø—	C(H)—Noncyclic groups
II	P—	Organoheteroid groups
	Q—	Rings containing no carbon
III	R—	Central atoms
	S—	Groups coordinated to P— or R—
	T—	Central atoms
	U—	Groups coordinated to P— or T—
	V—	Solvates
IV	Z—	Indeterminate structures
Unassigned	W—	
	X—	
	Y—	

* This includes fragments of heterocyclic rings containing a part of the group outside of the ring.

free elements, simple anions, and central elements of complex anions or neutral molecules (T). Assignment of free elements and central elements of neutral molecules to the R or T family is dependent on their electronegativity. Elements coordinated to the R and T families are coded in the

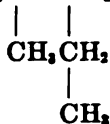
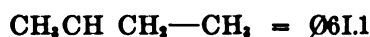
S and U families, respectively. The specific element and state of oxidation are described by the second and third characters in the group numbers. Solvate molecules are coded in Family V, while compounds of indeterminate structure are coded in Family Z.

Space has been left for future expansion in all sections of the code so that groups which may assume prominence at some later time may be accommodated.

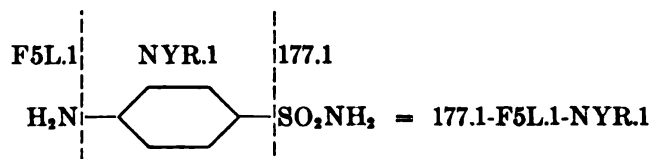
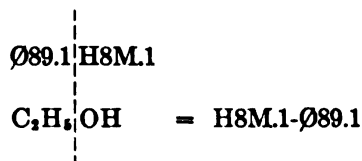
The limits of this chapter do not permit a detailed description of the specific rules covering the coding of compounds by the N.R.C. code. These have been prepared, and may be obtained by anyone interested*.

Very briefly, the procedure used to assign numerical designations to chemical compounds follows that described earlier in the discussion of the Frear code (pp. 468-70). Chemical compounds may contain groups which are classified in one or more of the divisions of families. For example, C_2H_5OH contains only organic groups, and both of these fall into Division I. $(C_2H_5)_3PO_4$, on the other hand, contains organic and inorganic groups which are classified in Divisions I and III, while $NaCl$ contains only inorganic groups which are classified in Division III.

Among the "organic" groups the degree of branching in noncyclic structures is not considered, but only the total number of carbon atoms linked directly together and the carbon to carbon unsaturation.

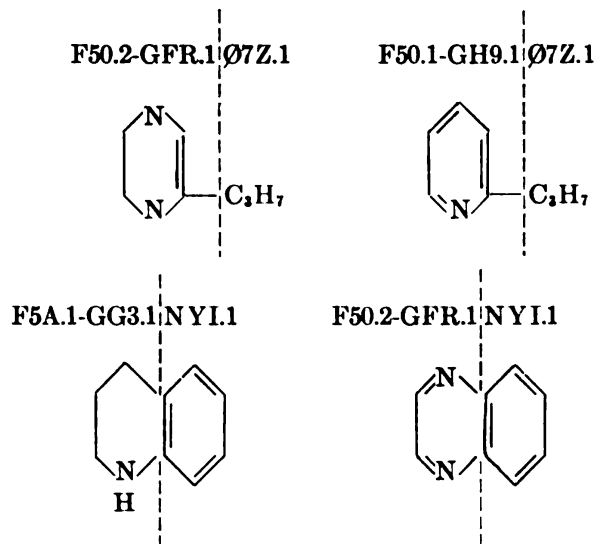


In breaking down compounds into separate groups for coding, the groups are ordinarily separated at the point of attachment to a carbon atom.

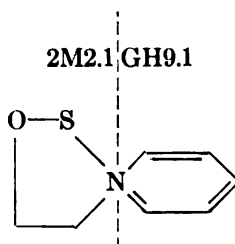


* Address inquiries to the Chemical-Biological Coordination Center, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

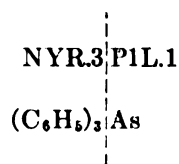
Separate designations are assigned to the several cyclic structures, taking into account the degrees of saturation. Heterocyclic structures fused to carbocyclic structures are separated for coding.



Spiro compounds are coded as two separate rings, with the common atom being counted in both rings.



Organoheteroid compounds have the heteroatom coded in Family P—.

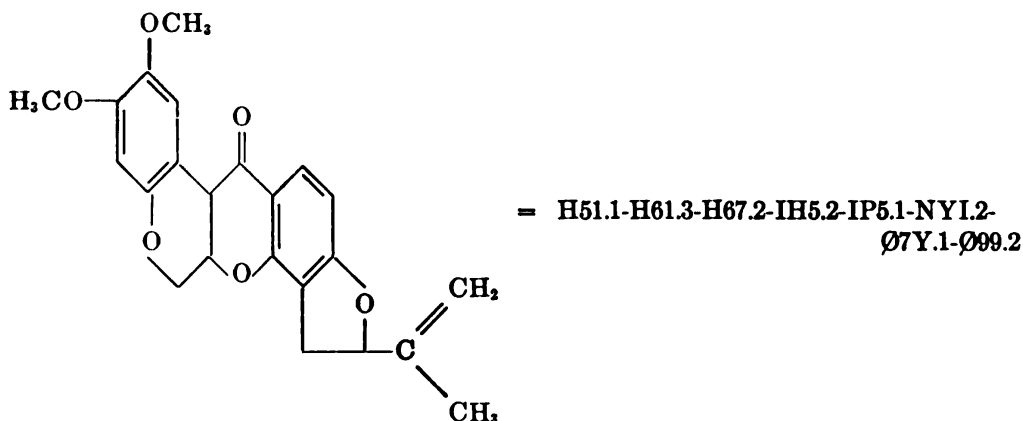


In ionic inorganic compounds simple combinations of two elements are coded by two group numbers, one indicating the element and its oxidation state present in the cation, another for the anion. In more complex structures one or more atoms in each portion is considered "central" and the other atom or atoms coordinated to it. For example, uranium UO_2^{+2} is RT2.1, sulfur in SO_4^{-2} is TNJ.1, etc. The oxygen atoms in UO_2^{+2} are coded in the S family as S63.2, while those in the SO_4^{-2} are U63.4. Thus, the combination of two group numbers, TNJ.1-U63.4, characterizes the SO_4^{-2} ion.

Solvate molecules associated but not chemically coordinated with organic or inorganic compounds are coded in Family V—. The $2\text{H}_2\text{O}$ in $(\text{Cr}(\text{H}_2\text{O})_4\text{Cl}_2)\text{Cl}\cdot 2\text{H}_2\text{O}$ is coded as V61.2, and the $4\text{CH}_3\text{OH}$ in $\text{CaCl}_2\cdot 4\text{CH}_3\text{OH}$ as V6A.4.

A few general examples of the NRC code designations are given below:

$\text{C}_8\text{H}_{10}\text{NCOOC}_2\text{H}_5$ 1-Piperidinecarboxylic acid, ethyl ester	= 630.1-GH2.1-Ø89.1-Ø99.1
$\text{CH}_3\text{CH}(\text{NH}_2)\text{CONHCH}_2\text{CONHCH}_2\text{COOH}$ Glycine, N-(N-alanyl)glycyl-	= 65C.2-F5M.1-H42.1-Ø7Z.1-Ø89.2
$\text{C}_6\text{H}_3(\text{NH}_2)_3\cdot 3\text{HCl}$ 1,3,5-Benzenetriamine, trihydrochloride	= F5L.3-NYR.1-RB6.1-T69.1



Rotenone

To summarize, the N.R.C. code has been devised for specific use with standard machine-sorted punched cards. It employs only arabic numerals and letters of the alphabet. The code number for any chemical compound is made up of a series of group numbers, each consisting of four characters, so that the constituent groups may be easily recognized even though they are not separated by signs. (Dashes and decimal points are used only for convenience, and are not recorded on the punched cards.) Constituent groups may be found easily by machine sorting for correlation studies. The code numbers may also be used for indexing and classification purposes by arranging them in numerical (and alphabetical) order.

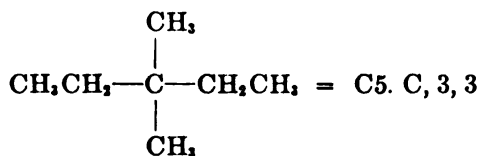
This system does not distinguish between position and stereoisomers, but this is not considered a serious fault since by elimination it is always possible to make a manual separation when only a few compounds are involved. It is true that the code designations are not unique, and occasionally two compounds having dissimilar structures will have the same code numbers. Although this may be a disadvantage, manual separation of a few compounds having the same code designations is a relatively simple matter.

Over 60,000 compounds have been coded by this system at the Chemical-Biological Coordination Center: most workers using the system have found it relatively simple and satisfactory to use.

The Dyson System^{11, 12, 13}

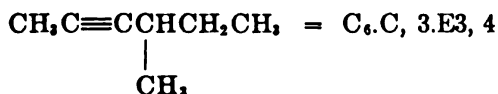
The Dyson cipher, probably the most ambitious of all proposed systems for coding chemical compounds, has been worked out in considerable detail for organic compounds. Like most other systems, it reduces the chemical structures to linear designations, composed of letters of the alphabet, numerals, and certain conventional symbols.

Carbon chains, which form the backbone of most organic compounds, are designated in the cipher in the same way as in usual chemical formulas. For example, ethane is designated as C₂, propane as C₃, decane as C₁₀, etc. If the structure is branched the cipher defines the longest chain first, then each side chain in turn, commencing with the largest. The points of attachment are indicated by whole numbers following the side-chain designation in the cipher.



The above indicates a five-carbon chain with two single carbons attached to it at the 3 position. The period between cipher designations indicates the completion of one operation. The whole numbers at the end of each operation (3, 3 in the example) are termed "locants".

Unsaturation in an acyclic residue is indicated by E for double bonds, and E3 for triple bonds.



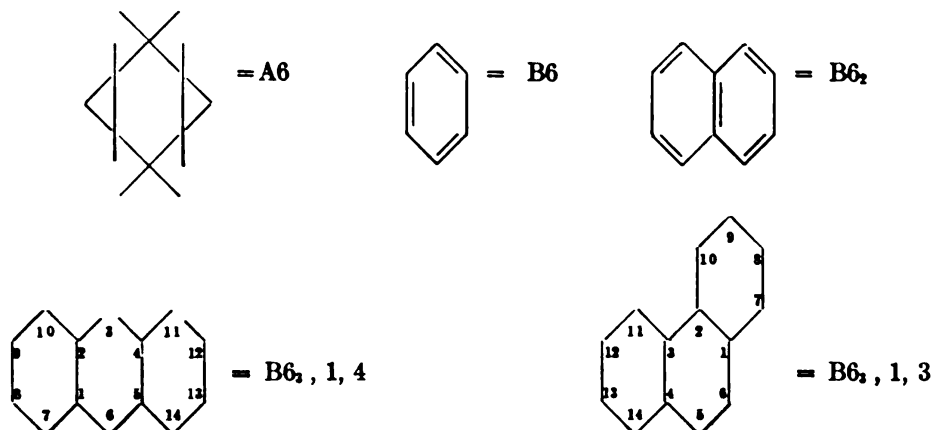
Cyclic compounds are designated as A (saturated) or B (aromatic) rings. Ring structures not completely saturated or aromatic are designated by the addition of the E symbol to indicate double bonds following the saturated designation, or H (for hydrogen) following the aromatic designation. The form requiring the fewest locants is chosen; or, if equal locants are to

¹¹ Dyson, G. M., "A New Notation and Enumeration System for Organic Compounds," 1st Ed., London, Longmans, Green, and Co., 1947.

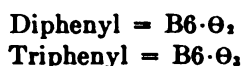
¹² Dyson, G. M., "A New Notation and Enumeration System for Organic Compounds," 2nd Ed., London, Longmans, Green, and Co., 1949.

¹³ Dyson, G. M., "A New Notation for Organic Chemistry," *Research*, **2**, 104-114 (1949).

be used, the aromatic (B) designation is used as a starting point. Fused rings are considered as being made up of individual single rings:

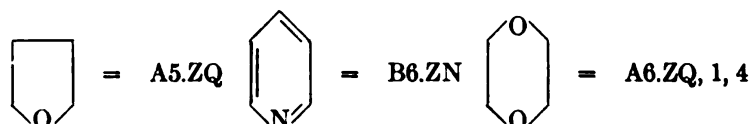


Coupled (not fused) ring systems containing two or more of the same kind of rings are indicated by Θ , to indicate repetition:

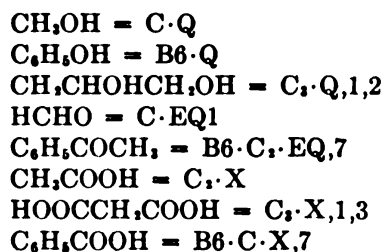


Positions of attachment may be indicated if it is necessary to do so to distinguish isomeric forms.

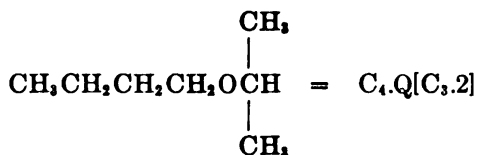
Heterocyclic structures are ciphered, using the symbol Z to indicate the presence of the heteroatom, followed by the symbol for the element. To avoid confusion, oxygen is designated as Q rather than O.



Certain conventions are followed for the more common functional groups. For example, while Q is used to designate oxygen in the hydroxyl group, EQ is used to indicate an aldo or keto group and the carboxyl group is ciphered as X.



Ethers or other compounds in which an atom other than carbon occurs in the chain are broken at the odd atom or atoms.



In this case the brackets indicate the second chain.

The sulfur analogs of oxygen groups are ciphered as S and ES; the nitrogen-containing groups have special designations to indicate nitro (N2), nitroso (N1), azo (N4), etc.

Space does not permit further discussion of the Dyson cipher, which, to be described completely, requires a book of some 132 pages. The basic principles enumerated above are elaborated and discussed in detail in this book, and any reader interested in coding chemical compounds is strongly advised to read it. The author claims that his cipher may be used with certain specially modified punched-card machines of the "Hollerith" type, and also states that it can be used with edge-punched cards, as well. However, it appears that sorting cards punched according to the Dyson system will require machines of special design, not commercially available at present. Whether these will become readily procurable at a reasonable cost is somewhat uncertain and this factor is one of the greatest practical drawbacks to the system.

Although its author has obviously spent a vast amount of time and effort in developing the Dyson cipher, it has a number of disadvantages and shortcomings. First, it appears to be extremely complicated, and it will require considerable time for a novice to become proficient in its use. The cipher for a complex compound tends to become unwieldy, and a number of conventions have been adopted to reduce the length of the cipher designation. Some of these are commendable, but in other cases the abbreviations used would confuse even a chemically trained worker. An example of this is C, 2(3)17, which indicates eight methyl groups attached in the 2, 5, 8, 11, 14, and 17 positions.

Probably the greatest weakness in the Dyson system is that it does not cover inorganic compounds. It has been the experience of the author that inorganic structures present many problems in coding, probably more in proportion to the number of compounds than in the organic field. Since any coding system should be applicable to all types of compounds, it appears that the Dyson cipher is only partially complete.

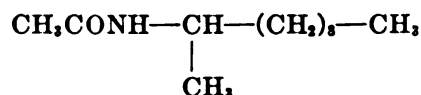
A number of other criticisms of Dyson's cipher have been voiced by various workers. In fairness to the system, it must be said that its author has done much to simplify and improve it¹³. The NRC volunteer testing program brought to light several difficulties which the author claims to have eliminated². Among other improvements, the symbol E has now been given a new meaning, and alphabetical citation of functional and other

non-hydrocarbon substituents adopted to conform to usage of *Chemical Abstracts*.

No specific information appears to be available at this writing as to the adaptability of the Dyson system to the newer computing machines. If such high-speed calculators can be used with the Dyson cipher, it may extend its usefulness.

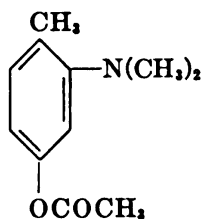
Gruber System¹⁴

In this system, aliphatic compounds are coded by numbers indicating the number of carbon atoms in the chain, plus designations for the attached groups: 1-dodecanol, for example, is C12.OH. The symbol Z is used to denote an interrupted carbon chain or hetero structure. To code the compound



Gruber uses the symbols Z13[C2(=O)NC10.4C]. The figure 13 refers to the total number of carbon atoms; the oxygen in the acetyl group is enclosed in parentheses to indicate its position as a substituent group. The methyl group is located by the designation 4C.

Gruber indicates a benzene ring by 6], so that the compound



is coded as 6]C·2Z2[N(C)C]4Z3[O(O=)C2], the dimethylamino structure being represented as Z2[N(C)C]. Likewise, the acetate group is coded as Z3[O(O=)C2.] Note that these groups of symbols are preceded by 2 and 4, respectively, indicating attachment to the benzene ring in these positions.

As in the other systems, special symbols and rules for fused rings and other complex structures are given by the author. In the trial coding tests conducted by the NRC, it was noted that some difficulty was encountered in determining the order of citation, and in enumerating the rings. More recently Gruber has contributed a number of suggestions to the Wiswesser system, and these have been incorporated into the latter. As far as is known, the Gruber system has not been used to any extent, either in this country or abroad.

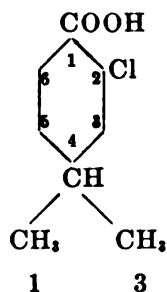
¹⁴ Gruber, W., Die Genfer Nomenklatur in Chiffren und Vorschläge für ihre Erweiterung auf Ringverbindungen. *Angew. Chem. Beiheft*, 58 (1950).

Opler and Norton (Dow) System^{15, 16}

The basis of this coding system is actually one used for encoding networks. Chemical compounds are regarded as networks of chemical groups. The authors have selected 332 structural elements, such as ether, ester, etc., and to each of these is assigned an arbitrary number of three digits. The mode of attachment of these groups one to another is indicated by three additional digits, and a seventh digit serves to number the group. No special order of listing the groups is necessary. Thus each structural element or group in the compound is represented by a group of seven digits, and the complete structural representation consists of as many of these seven-digit units as may be required to characterize the compound.

As a general example, let A represent the structural element (or group) being considered in compound AB. *The first digit* in the seven-digit unit gives the location in some other groups (B) to which the group (A) being coded is attached. It is called the locant. *The second digit* tells by which of its positions group (A) is attached to group (B). *The third, fourth and fifth digits* designate the group number assigned to group (A). *The sixth digit* is the identifying number of the group (B) to which group (A) is attached, and *the seventh digit* is the identifying number of the now coded group (A).

Specifically, to code 2-chloro-4-isopropylbenzoic acid,



the compound is divided into four groups as follows:

1. Benzene ring
2. Chloro (—Cl)
3. Acid (—COOH)
4. Propyl (—C₃H₇)

To code, list the three-digit numbers assigned to these groups (enter as third, fourth and fifth digits in the seven-digit unit).

¹⁵ Norton, T. R., and Opler, A., "A Manual for Coding Organic Compounds for use with a Mechanized Searching System." Published by the Research Dept., Western Division, Dow Chemical Co., 55 pp. (March 15, 1956).

¹⁶ Opler, A., and Norton, T. R., New Speed to Structural Searches, *Chem. Eng. News*, 34, 2812-16 (1956).

—	—	1	0	6	—	—	benzene
—	—	0	6	1	—	—	chloro
—	—	3	0	4	—	—	acid
—	—	0	0	3	—	—	propyl

Next, number the groups sequentially, starting with any group (digit 7).

—	—	1	0	6	—	1	benzene
—	—	0	6	1	—	2	chloro
—	—	3	0	4	—	3	acid
—	—	0	0	3	—	4	propyl

For each group, list the number of the previous group to which it is attached. Benzene, having group 1 in this example, has no number under this step. The previous group to which chloro is attached is benzene group 1, etc. This digit is placed in the sixth position:

—	—	1	0	6	—	1	benzene
—	—	0	6	1	1	2	chloro
—	—	3	0	4	1	3	acid
—	—	0	0	3	1	4	propyl

In the second position of the seven-digit number list the position in each group by which it is attached to the previous group. Propyl is attached to the benzene at its 2 position to benzene.

—	—	1	0	6	—	1	benzene
—	1	0	6	1	1	2	chloro
—	1	3	0	4	1	3	acid
—	2	0	0	3	1	4	propyl

The first digit indicates the position in the previous group to which the group being coded is attached. For example, chloro is attached to the 2 position on the benzene ring: note that benzene again has no number in this step.

—	—	1	0	6	—	1	benzene
2	1	0	6	1	1	2	chloro
1	1	3	0	4	1	3	acid
4	2	0	0	3	1	4	propyl

The completed code for 2-chloro-4-isopropylbenzoic acid thus becomes
/ — — 106 — 1/2106112/1130413/4200314/

Certain conventions are followed, such as the use of the figure 9 in the

first position of the seven-digit code number to indicate indeterminate structures; bridged, spiro and fused ring compounds are identified with a two-digit (instead of a normal three-digit) number in the fourth and fifth positions. This number begins with 7, so that the figure 7 occurring in the fourth position automatically indicates this type of structure. Stereo- and optical isomers are indicated by numbers between 900 and 999.

The advantages claimed for this system are its simplicity and adaptability to high-speed computing machinery. It is reported that coding compounds of average complexity required less than two minutes per compound. Once the compounds have been coded—reduced to their digital equivalents—and key-punched on cards, they are ready for storage in and search by a high speed computer. The searching principles are the same, whatever machine is used; only the specific instructions fed to the machine are different. Experience has indicated that punched card machines alone were not satisfactory for searching these codes. Data processing machines, capable of extremely rapid operation, such as the IBM 701, were satisfactory, since such a machine will perform approximately one million simple operations per minute. Opler and Norton are of the opinion that the system they have devised, or some improved modification thereof, will be widely useful for high-speed structural searches. With more and more computers being built, such a system will be available generally in a short time. Opler and Norton¹⁷ have prepared a manual for programming computers for use with this system. This includes a list of characteristics required in a satisfactory computer, as well as detailed directions for carrying out the searching operation.

This system, when used with high-speed computers, appears to be well adapted to studies of organic compounds. The authors have not included inorganic compounds in their scheme, but adaptation to include them should not be difficult. The high cost of the electronic computing machines and their present restricted distribution may prevent some workers from using this system.

Wiswesser Notation System^{18, 19}

Wiswesser's "line-formula" notation, first proposed in 1950, resulted from a ten-year series of attempts to describe complex ring structures in a simple yet logical manner, and to calculate hydrocarbon isomers in terms of familiar mathematical series. In 1945 F. D. Rossini and his associates²⁰

¹⁷ Opler, A., and Norton, T. R., "A Manual for Programming Computers for Use with a Mechanized System for Searching Organic Compounds". Published by the Research Dept., Western Division, Dow Chemical Co., 23 pp., (April 25, 1956).

¹⁸ Wiswesser, W. J., "A Line-Formula Chemical Notation," T. Y. Crowell Co., New York, 1955.

¹⁹ Wiswesser, W. J., The Wiswesser Line Formula Notation. *Chem. Eng. News*, 30, 3523-26 (1952).

²⁰ F. D. Rossini *et al.*, *Bur. Stand. J. Res.*, 34, 413-34 (1945).

at the National Bureau of Standards published a report showing that hydrocarbon properties could be calculated precisely only in terms of the primary, secondary, ternary, and quaternary carbon atom components. These and other physical property correlations with structure (e.g., characteristic infrared frequencies) convinced Wiswesser in 1950 that the basically significant hydrocarbon segments are not the "carbon skeleton" abstractions of the Geneva *nomenclature*, but the pictorially obvious alkyl chains and branched carbon atoms of the much older "line-formula" *notation*. Thus Wiswesser solved the isomer-calculating problem in 1955 through this natural "branch group" analysis of structures. Much earlier clues on notational design were found in statistical analyses of large chemical catalogs—starting with Frear's functional group frequency data of 1946—and in a revealing historical study of structure-printing traditions (reported to the ACS Meeting in Chicago, 1950).

Mnemonic aids were used extensively in a deliberate effort to create a system of symbols that would be the simplest to learn and easiest to use of all printable alternatives. Thus only eleven new letter symbols are required for the chemical groups, as explained below. An unexpected bonus from this straight-forward "least effort" analysis of the problem is that the resulting expressions also yield the most concise of all proposed notations, and lend themselves readily to profitable applications with IBM or similar tabulating equipment.

The first characteristic feature of this notation is that the atomic symbols are cited "end to end," in a pictorially direct connecting order (Rule 1). This practice follows the oldest structure-delineating tradition, and eliminates much unnecessary "enumerating" labor. Thus combinations of familiar symbols produce instantly recognizable descriptions, even when the structures themselves may be unfamiliar—as in NCSCN, OCCCCO, or SCCCCS.

The second characteristic feature is that all otherwise equal sequences of symbols are resolved through their self-evident alphabetic order (Rule 2). Preference always is given to the *highest*-ranking sequence of open-chain symbols, in order to preserve the traditional emphasis on terminal functions rather than the carbon-chain symbols. Thus asymmetric combinations of familiar symbols are resolved as shown here: IF, NCF, OC, ONN, and SCO.

Arabic numerals in the Wiswesser notation are reserved to denote the number of carbon atoms in *normal* alkyl groups or in unbranched polymethylene segments (and in the corresponding cycloalkyl rings). Thus the first five *normal* alkanes are denoted 1H, 2H, 3H, 4H, and 5H.

Punctuation marks show modes of connection or disconnection, such as the branch-terminating period mark. Thus butyl ethyl propyl methane is 4Y3.2. The colon denotes a single unsaturation in the familiar manner of use, and a double-colon mark denotes the doubly unsaturated acetylenic

link, with quantum-mechanical accuracy as a "double double bond." Thus ethylene is 1:1 and acetylene is 1::1.

Capital letters are reserved in the traditional manner to denote specific atomic groups. Thus Y and X denote the physically significant Y-branched and X-branched aliphatic carbon atoms; E and G (both from haloGEN) denote the two very common halogen atoms—bromine and chlorine. Thus E5 is 1-bromopentane and GYGG is chloroform.

Oxygen and nitrogen functions outnumber all other functions by a huge prominence; therefore six of the remaining seven special letter symbols are assigned as shown:

Z for the terminal NH ₂ -group (from hydraZine)	Q for the OH-group (from aQua)
M for the NH-group (a "Mid-aMino")	V for the —CO— connective
K for the quaternary N-atom (“kationic”)	W for the “double-O” in —NO ₂ and —SO ₂ — groups

Thus 2M2 is diethylamine, WNQ nitric acid, ZQ hydroxylamine, and ZVZ urea.

The last special letter symbol also is reserved for a cyclic group of enormous prominence among known structures—the benzene ring. Simple benzene derivatives are numerous enough to justify a special classification within the carbocyclic compounds; accordingly, this special letter R (for resonating, regular-hexagonal ring) is given last and lowest rank in the Z, Y, X, . . . C, B, A, 9, 8, 7, . . . 2, 1, R sequence of atomic group symbols. Thus 1R is toluene, 2R is ethyl-benzene, QR is phenol, RR is biphenyl, and ZR is aniline.

Lower case letters are used in a distinctive manner (first suggested by Kekulé in 1866) to locate relative ring positions. The first two, a- and b-positions need not be specified when the corresponding atomic group symbols are next to the benzene ring symbol. Thus the *ortho*-phenylene connective is denoted —R—, the *meta*- is —Rc—, and the *para*- is —Rd—. Aspirin is QVROV1, resorcinol is QRcQ, and benzidine is ZRdRdZ.

Branched benzene derivatives and open-chain structures both are delineated by a third familiar rule of procedure: at branched points, first cite the side group having the fewest coding symbols. Thus the longest chain of symbols is depicted as the “main line” that begins and ends the atomic delineation.

Numerals have been used for more than half a century to denote the number of atoms in a given ring—an *absolute* measure. These same ring numerals are used in this notation, distinctively enclosed in parentheses along with the other heteratomic or heterocyclic symbols that are a part of the ring description. The ring numerals are punctuated with a stroke mark if

the ring is saturated or "nonaromatic." Thus (3/) is cyclopropane and (6/) is cyclohexane; (5/M) is pyrrolidine and (6/M) is piperidine. Heterocyclic numerals are punctuated with a period mark if the ring is dehydrogenated to the aromatic limit; thus (5.O) is furane, (5.S) is thiophene, and (6.N) is pyridine.

Within the ring parentheses, all otherwise equal delineations are resolved by selecting the sequence that gives the *lowest* possible measure of every kind—lowest set of heterocyclic symbols, lowest set of locants, and the like. Thus pyrazole is (5.MN) and thiazole is (5.N cS). Next, the ring branches are cited in simple alphabetic order of their locants. Thus proline is (5/M) bVQ and niacin is (6.N) cVQ.

Bicyclic systems can be pictured very readily through the use of prime marks that are inserted between the ring numerals to show the atomic bridges. No "saturation stroke" is necessary with prime marks, since bridged ring structures are theoretically nonaromatic. Thus norpinane is (4'6) and quinuclidine is (6''6/cN). The stroke mark is added to show saturation in unbridged systems, as with (6/6/) for decalin and (66/) for tetralin. Aromatic character is implied in the (66) notation for naphthalene, and the (66.bN) notation for quinoline.

All polycyclic ring positions can be determined by a single atom-to-atom lettering procedure: start the longest possible chain of ring positions at the point that gives the lowest sum for the lowest locants in each ring. Thus in fused and bridged systems, the position-determining chain starts at a fusion atom or a bridge atom.

Summarizing the Wiswesser Notation System, it can be said that this is one of the simplest systems to be proposed. The NRC survey, discussed earlier, found that this system required less time to encode and decode compounds than the Dyson, Gruber, and Silk systems. It was also found, however, that the percentages of disagreement when using the Wiswesser notation were higher (both for encoding and decoding) than these other systems. The chief criticisms of the Wiswesser code involved certain rules which appeared to be unclear or unspecific. These, the author indicates, have been corrected and the system now appears to be both simple and workable.

Berry-Crane Composite System

The testing of several coding systems by the volunteer group described previously in this chapter² resulted in no unanimous conclusions, but it was the general feeling that no one system yet proposed was worthy of adoption as an international standard. Berry and Crane^{2, 3} have proposed a composite system embodying the best features of several other codes, and employing redundancy, or paraphrasing of the code.

Basically, the Berry-Crane notations consist of two parts. The first (or code) part makes certain general statements about important structural features of the molecule, such as the number and kind of rings, if any, the number and kind of functional groups, and the number and kind of elements present. The second part of the notation (the cipher) is a complete and independent set of unambiguous instructions for reconstructing the molecule. Some information which could be left to inference is explicitly stated.

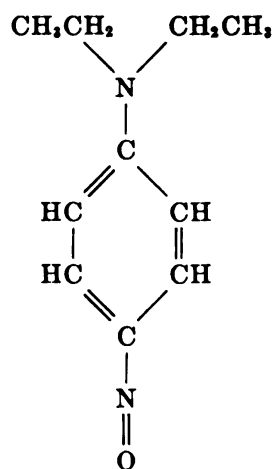
In the code section, the functional groups have been sorted into 26 classes. A fully substituted carbon atom is always C; the CH group is J; CH₂ is L; and CH₃ is E. The symbol G represents a carbonyl group; a colon represents a double bond; a semicolon is a *cis* double bond, and an exclamation point a *trans* double bond, for example.

To translate the formula for N,N-diethyl-*p*-nitrosoaniline into the Berry-Crane system, it is stated that one ring is present; no fused rings, no hetero rings, and the ring contains six members—this information is represented by the number 1006. The fact that the ring is aromatic is indicated by the symbol R, and the four double bonds by the figure 4. A comma is inserted to indicate the end of this section. The code part of the designation continues with the letter M, indicating ammonia, amine, or amine salt group; T is used to designate nitrous acid, nitrite or nitroso compound. A second comma precedes the molecular empirical formula index, which is cited in the order used in *Chemical Abstracts*, with spaces between the two-digit numbers, and followed by a period. The complete code section of the representation for N,N-diethyl-*p*-nitrosoaniline thus becomes 1006R4, MT, C 10 H 14 N2O.

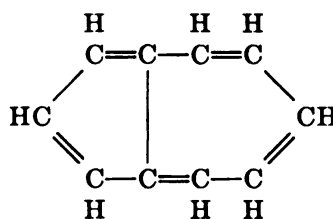
The cipher further describes the structure of N,N-diethyl-*p*-nitrosoaniline as follows: the substituted amino group is symbolized as E (for CH₃) L (for CH₂) NLE. The period indicates that this completes the description of this substituent group. The point of attachment is arbitrarily shown as a, while the benzene ring is (J6), and in the d position on the benzene ring is a N:O group. The complete cipher, then, becomes ELNLE.-a(J6)dN:O. Combined with the "code," the complete designation for N,N-diethyl-*p*-nitrosoaniline becomes 1006 R4, MT, C 10 H 14 N2O. ELNLE. a(J6)dN:O.

Azulene (see following formula) has two rings, two fused rings, no hetero rings, and one five- and one seven-membered ring; hence 22057, etc.

The Berry-Crane system is frankly a composite of the better features of several other codification methods. The experience and suggestions that came from the IUPAC Commission deliberations and the NRC volunteer testing program have been utilized in the Berry-Crane system, hence it should avoid many of the pitfalls with which the others were plagued. Up



1006R4, MT, C 10 H 14 N2O.
ELNLEa(J6)dN:O
N,N-Diethyl-*p*-nitrosoaniline



22057R5, C 10 H8.
(J5)ab(J7).
Azulene

to the time this chapter was written, no extensive test of this system has been made, although one is reported to be planned. The results of this trial should demonstrate conclusively the usefulness of this system.

The Code of Gordon, Kendall, and Davison²¹

The basic concept of this code is the chemical species, which is defined as "a set of atoms individually given, given pairs of these atoms being linked together by directed bonds. The net charge of the set and of any discrete ion contained in it must be specified." Thus, no distinction is made between the various types of chemical bonds.

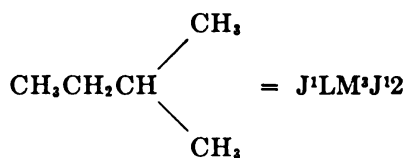
Commonly used chemical symbols are employed for the elements, with the following additions:

- (1) CH₃, CH₂, and CH are represented by letters J, L, and M, respectively.
- (2) X is used to indicate ring closure.
- (3) E and G indicate negative or positive net charges.
- (4) Isotopes are indicated by the letter T.
- (5) Repeating (polymeric) units are indicated by Q.
- (6) Structures designated for purposes of classification are indicated by R.
- (7) Discrete portions of the molecule are separated by oblique lines (/).

In translating chemical structure to the cipher the symbols are listed in sequence as the elements occur in the compound. Certain rules for the seniority of symbols are stated so that all workers will begin the cipher at

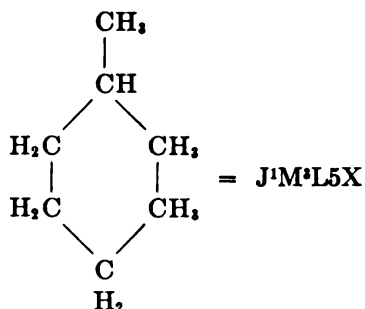
²¹ Gordon, M., Kendall, C. E., and Davison, W. H. T., "Chemical Ciphering," London, The Royal Institute of Chemistry, 1948.

the same point. To the symbols representing the elements are added superscript numbers indicating their co-ordination numbers. Branch points in carbon chains are thus indicated:

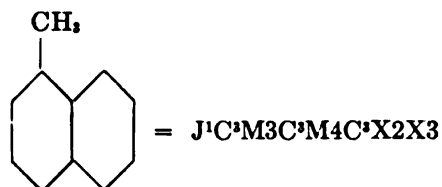
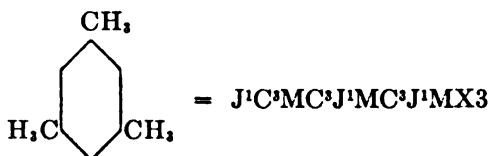
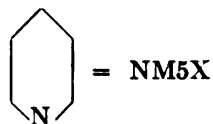


The choice of a starting point is covered in the rules and varies, of course, with the type of compound and the terminal groups present. Briefly, symbols of lower co-ordination numbers take precedence; while in symbols having equal co-ordination numbers the order is elements other than carbon (in order of atomic number with the lowest first), hydrocarbon groups J, L, and M, and carbon.

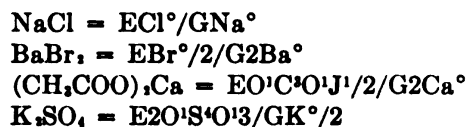
Ring compounds are ciphered using the symbol X to indicate the point of closure. For example, methyl cyclohexane



Other examples are as follows:



Ionic compounds are ciphered, using the appropriate symbols E and G mentioned earlier. For example,



Several special rules cover isotopes, compounds of indefinite composition, and macromolecules. Provision is also made for indexing ciphers so that they may be used for classification purposes.

In summary, the Gordon, Kendall, and Davison cipher appears to have the advantage over the Dyson system (cf., page 476) in its relative simplicity. It seems to be fairly easy to learn, but it is extremely doubtful whether a chemist could be trained in its use in the short time claimed by the authors. Furthermore, it is not likely that it will be used by nonchemists. Certainly the average individual without chemical training would find it difficult, if not impossible, to understand the rules, to say nothing of their application.

This system can be adapted readily to computers²², but some provision must be made to enable such machines to distinguish bonds, specific atoms, etc., rather than merely to scan the codes. With such adaptation it should be possible to pick out functional groups or other desired structural features.

U. S. Patent Office System

A system for coding chemical structures is being developed at the U. S. Patent Office, and an abbreviated description has been published²³.

Only general comments will be made here concerning the system, since a fuller description of it is given in the chapter on the research activities of the Patent Office (Chapter 12). Quoting the published report, "The method of coding takes cognizance of each element present in the chemical compound and its graphic structural relationship to every other element so that any selected fragment of the entire compound can be recognized and retrieved by machine when a search is made for the class of compounds containing that fragment. In addition, the complete configuration gives a code uniquely different from the code for any other compound."

As a simple example of how the Patent Office system works, the com-

²² Davison, W. H. T., "Programs and Equipment for Sorting Gordon-Kendall-Davison-Punched Cards for any Structurally Defined Groups." Paper presented before the Division of Chemical Literature, American Chemical Society Meeting, September, 1951.

²³ Lanham, B. E., Liebowitz, J., and Koller, H. R., "Advances in Mechanization of Patent Searching—Chemical Field." Patent Office Research and Development Reports, April 11, 1956.

pound $\text{HOCH}_2\text{CH}_2\text{OCH}_2\text{CH}_2\text{Br}$ is coded as Q H-O-C-C-O-C-C-Br Q, the symbol Q being a code designation for grouping signals indicating a chain. Thus in a search for an ether, the combination Q-C-O-C-Q will be sought; for an ether-alcohol in which the two oxygens are in a 1-4 relationship will be sought as Q-H-O-C-C-O-C-Q. More complicated ring structures, heterocyclic compounds, etc., are represented by rather lengthy ciphers.

The Patent Office research team is simultaneously developing a system of coding for structure searching which employs a topological approach, and in which a computer will be employed for searching. This system is also more fully described in the chapter on the Patent Office research program (Chapter 12).

The Patent Office structure coding systems were developed for searching purposes only, and were not intended to be used for indexing.

Chodosch System

Another system employing a topological approach to coding structures is being developed by Robert Chodosch, at the University of Florida. The technique is applicable to encoding networks in general. Since the system has not yet been tested on a sizeable number of compounds, it will be discussed only briefly here*. The process of encoding involves (1) labeling the points in a structure; (2) preparing a table in which points connected to each other are listed in columns; and (3) substituting integers for the points and rearranging the resulting tables to produce the lowest possible integer table. This is the unique notation for a given network. Indexing the tables can be done dictionary style, following particular rules for ordering the lowest tables. Structure searching can be accomplished by scanning the stored integer tables for any which contain the desired network. Overlapping of tables will be found whenever a compound contains the particular structure being sought. The system has been designed to be used with a computer which can calculate the lowest possible integer tables, and scan for overlapping portions of tables in the searching process.

Zatopleg System²⁴

This system, devised by Mooers, has been described only briefly by the author. The procedure used to encode a compound includes: (1) drawing a structural formula showing the position of every atom in the lowest energy state of the molecule; (2) in any arbitrary (or random) order, numbering every atom in the formula serially, 1, 2, 3, etc.; (3) listing of the numbers

* Further inquiries should be addressed to Robert Chodosch, Rutgers University, P.O. Box 821, New Brunswick, N. J.

²⁴ Mooers, C. N., "Ciphering Structural Formulas—The Zatopleg System." Published by the Zator Co., Boston, Mass. (1951).

assigned to carbon atoms, and so on for all types; and (4) listing of the number-pairs of atoms that are joined with a single chemical bond, in another list those joined by a double bond, etc. It is claimed that the numerical representation obtained by this procedure can be mechanized by an electronic computer, such as UNIVAC. However, as far as can be determined, no large-scale trial of Zatopleg has ever been carried out.

Nodal Index for Branch Structures

H. P. Luhn, of the IBM Research Laboratories, has developed a system of notation which can be used to record chemical structures, the flow of processes, or the assembly of mechanical and electrical circuit elements²⁵. The system is based on a statistical method of serial recording and later serial analysis. For a given structure there is derived a unique expression consisting of a set of topological descriptions of portions of the structure. These portions are called nodes, and are overlapped so that at least two elements are common to a pair of adjoining nodes. Each node is treated as an independent entity so that the order in which nodes are enumerated does not indicate topological relationships. Instead the nodes are given in a particular order, which may be established to express some useful characteristics and which becomes part of the rules. The order may be, for example, the valence of chemical elements. A similar order is to be assigned to the elements within each node.

Structures are compared to determine whether two structures are similar, or whether a given structure is contained wholly or in part in another structure. The comparison can be carried out by trying to subtract the first notation from the second. If the subtraction can be completed the first structure is probably fully contained in the second. If it cannot be subtracted, the first structure is not included in the second. An index may be established listing derived notations in the basic order of the system together with their pictorial counterpart. As far as is known, the system has not yet been tested on a sizeable number of compounds.

The task of devising a notational system for chemical compounds is an extremely difficult one. It might well prove to be impossible to devise any one system which will accommodate the varied types of both organic and inorganic structures, for solutions of varied problems. Simpler types of codes may be satisfactory for particular types of investigations; on the other hand, use of versatile searching equipment may call for development of more sophisticated coding schemes. Extended use under practical conditions is the best way to determine the suitability of a particular type of code to a particular problem.

²⁵ Luhn, H. P., A Serial Notation for Describing the Topology of Multi-dimensional Branched Structures. (Nodal Index for Branched Structures). International Business Machines Corporation, New York (1955).

Chapter 23

SUPERIMPOSED CODING WITH THE AID OF RANDOMIZING SQUARES FOR USE IN MECHANICAL INFORMATION SEARCHING SYSTEMS

H. P. LUHN

*IBM Research Center
Yorktown Heights, N. Y.*

Introduction

The mechanical process of scanning records for the purpose of selecting those which contain wanted information has presented the problem of how to record such information most effectively. The type of information referred to here is unlike that used on business records where a limited number of classes of information terms are used. Because of the limited number of such terms, it is feasible and has been the custom to provide fixed areas or fields on records wherein the appropriate information can be inserted.

With the use of punched cards, this system of fixed fields became particularly significant because once a card reading machine was adjusted for a given record format, there was no question as to the meaning of the information recorded therein.

When information of a more general character is being recorded, great difficulty is encountered in assigning fields to the many classes of information terms that might possibly occur. In the first place, the number of such classes may vary from record to record and, in the second place, there may be more than one term that can be assigned to a given class. Obviously, in allowing for a maximum of such variations, the number of fields would have to be so numerous that the record form would assume impractical dimensions.

One way of overcoming this obstacle is to abandon the concept of fixed fields and, instead, to record information in serial form, separating and identifying classes of terms by special division marks. This method has been used in experimental column-by-column card scanning systems¹ and in systems where continuous tapes serve as recording means.

¹ H. P. Luhn, "The IBM Electronic Information Searching System," May (1952), IBM Research Center, Yorktown Heights, N. Y.

In cases where the above type of recordings are not feasible, two systems have been developed to give the effect of serial scanning.

The first system, using punched cards in conventional card processing devices, consists of duplicating a given record as many times as there are terms to be scanned in such a way that each of the resulting cards has a different one of the terms in a fixed field. This process is sometimes referred to as "field rotation" and the effect is that after reading the fixed field of all the cards in the set, all the information contained in the record has been scanned.²

The second method, designed to bring about similar results with the use of a single card, is referred to as "Superimposed Coding" and consists of recording a plurality of information terms, one over the other, into one common field. The merging of such codes necessarily produces secondary combinations which might represent unintended, yet valid terms. It is therefore necessary to provide means which will minimize the interference caused by such spurious information.

In solving this problem, reliance is made on the randomness in which letters or numbers happen to be combined to constitute a term and in which such terms are being used. If such randomness is substantially absent among a set of given terms, a re-coding method may be reverted to. One method consists of substituting random numbers for the given terms (Chapter 10).

The objective of the system, which will be described below, is to achieve the effect of randomizing non-random terms without re-coding them.

A New Scheme of Superimposed Coding

The superimposed coding schemes described here are based on word coding to the extent that a pair of letters is recorded as a single mark within a two-dimensional recording area. In adapting it to punched card operations, the peculiarities of punched card equipment and processing machinery have been taken into consideration. The most important peculiarity is that most standard machines are designed to read cards in a parallel fashion. In the case of an IBM card all of its 80 columns are read simultaneously, that is, in parallel. The individual marks within the 12 possible positions in each column are read serially on a differential time basis. Thus, a given hole is identified by its column and by the instant (in time) at which it passes the reading elements of the machine.

Superimposed coding schemes rely on this two-dimensional arrangement of recording and are identified as the intersections of columns and rows within a field of a fixed size. When information recorded in this fashion

² An example in point is: The Punched Card System of the Chemical-Biological Coordination Center, National Research Council, Washington, D. C.

has to be read for the purpose of scanning, the process of comparison or matching has to be done in this parallel and serial fashion. This process is wasteful in time and equipment as well as recording space utilization. The new scheme, therefore, proposes to rearrange the contents of a two-dimensional field and to record it in one-dimensional form in a single row across a card. The effect of this is that the equivalent of 12 such fields may be read consecutively with the passage of a single card. The advantages of this approach will be briefly described.

In recording information for searching purposes, it is desirable to signify the relationship of the various information elements. Records enumerating a number of things together with their characteristics should reflect which characteristics refer to which thing. Superimposed coding does not permit such relations and differentiations to be expressed in a single field. The remedy might consist of using a plurality of fields, with each one containing information elements that are directly related. However, the provision of several fields side by side on a single card would defeat its basic simplicity and would require special equipment within the machine for multiplexing the process of scanning. Using as many cards as there are fields required, on the other hand, would not only increase the size of files and the searching time, but would also have many other drawbacks.

In the art of information searching the terms "words" and "sentences" are often used to express the relationship between the various information elements. Words within a sentence express a closer degree of relationship to each other than words in different sentences. The information elements punched into a given field may therefore be referred to as the "words" and the total of these words within a field may be referred to as a "sentence." In using the linear form of superimposed coding as proposed in the new scheme, as many as 12 sentences may be punched on a single card. Each of these sentences is read in a parallel fashion for a single-cycle comparison or matching operation. If desirable, several sentences in sequence may be tied together by special marks to form the equivalent of "paragraphs."

Construction of the Code

As indicated earlier, superimposed coding produces a certain amount of unwanted, though valid combinations or words. When constructing such codes, special attention is therefore directed to minimizing such spurious words or the effects caused by their presence. The quality of resolution of such a scheme depends on a number of variables such as:

1. Size of collection of records.
2. Size of the recording field.
3. Number of recording fields.

4. Number of marks used per word.
5. Number of words constituting the dictionary.
6. Number of words entered into a field.
7. Number of words to be matched.
8. Degree of randomness of the letters or numbers employed in the words.

The statistical aspects of superimposed coding schemes have been investigated and described by C. S. Wise and others (Chapter 21).

In dimensioning and constructing the new scheme, proper recognition has been given to the findings of the above authors. It should be realized, however, that at this time there is not available any statistical information derived from actual applications which might confirm the theoretically derived values for spurious matches. Furthermore, the degree of tolerance of a system for unwanted selections may differ with the field of application. Whatever these values may be, the system is "failsafe" in that it produces at least all of the required matches.

Another feature that has to be considered in return for the compactness derived from superimposed coding is the fact that once information has been thus encoded and superimposed within a field, there is no obvious way of decoding it back into the words originally encoded. It is therefore necessary to list these words in a more conventional manner on the respective coded records or to maintain a master file which can be referred to by way of a reference number when it is desired to identify the words actually encoded.

As pointed out earlier, word coding is being employed in most cases to overcome restrictions and difficulties imposed by information-handling facilities. The process of encoding and decoding words with the aid of code books or dictionaries is time-consuming, and any scheme which will simplify this task will be a desirable improvement. If, for instance, words could be spelled out directly without the aid of a code book, a great deal of time and effort could be saved. The new scheme pays particular attention to this phase, and a method has been derived which makes such a procedure reasonably feasible.

Because this aspect of a system is so desirable, the new encoding scheme will be developed by applying it first to words in their original spelling. Let us assume a square or matrix having 26 rows and 26 columns. By writing the 26 letters of the English alphabet along both coordinates, all two-letter permutations of the alphabet are designated by the 676 intersections. Now, considering the horizontal axis to represent the first letter of a pair and the vertical axis the second letter of a pair any word may be spelled out in the following manner. Take the word "CHESTER." The first pair of letters is CH, which is represented by a mark at the intersec-

tion of the C row and the H column. Subsequent letter pairs of the word might be marked similarly at the appropriate intersection of the matrix. If an odd letter remains at the end of a word it can be paired with a blind symbol represented by a 27th column. The result would be the entry of four marks into the matrix.

Instead of operating in this manner, however, it is a feature of the new method to spell progressive pairs in single letter steps. In this manner, the word "CHESTER" would be spelled as 7 pairs, namely CH-HE-ES-ST-TE-ER. Thus an interlinked chain is formed in which the second letter of each pair is also the first letter of the immediately succeeding pair. This interlinking may be carried out one step further by closing the chain, in that the last letter and the first letter of a word are considered as the final pair. Therefore, the complete spelling of the above word is CH-HE-ES-ST-TE-ER-RC and the number of resulting marks equals the number of letters in the word.

The result of this method of chain and ring spelling is that the sequence of letters has been established as a closed system and any mark outside this system or ring must belong to another word. Furthermore, the end-around spelling of the last and first letter, namely the pair RC, prevents a match with a portion of a word like "ROCHESTER," which would differ in the end-around spelling of RR. By the same token, a word like "CHEST" would not match with portions of either of the two previous words because of the spelling of the end-around pair as TC. The occurrence of the word "CHEST" in CHESTER and ROCHESTER may, however, be ascertained by searching for this word minus the end-around link TC. It is apparent that if the words had been spelled in the form of unrelated pairs of letters, such differentiations would not have been possible.

When the words are spelled as a ring, there may be a question as to where a word begins. If it is important to indicate this, the addition of an extra letter such as Q at the end or the beginning of the word could serve to mark the break in the ring. Another way of marking the end of a word would be to omit the end-around spelling and, instead, pairing the last letter with an 'End' symbol represented by a 27th column.

Additional words may be similarly spelled and added in this square. Intersections of the resultant chains would, of course, remove the possibility of unique interpretations of the marks. These points of confusion are, however, less given to dilution than a system wherein the various marks are entirely unrelated.

While the 26 by 26 square might be desirable from a safety point of view, its size is impractical and actually wasteful. Usually the size of a square may be reduced substantially without seriously impairing its usefulness because of the following considerations.

The statistical rate of usage of the letters of the alphabet in spelling words varies considerably. Also, certain letter combinations may never occur. The 26 by 26 square may therefore contain intersections which will never be used. Others will be used rarely and still others will be used most of the time. The objective therefore is to design a scheme where the probability of a mark appearing in any one of the fields is reasonably even. As mentioned earlier, replacing the words with random numbers accomplished this, but the feature of the new system was to avoid this type of re-coding.

Randomizing the marks is accomplished instead by reducing the size of square and by assigning several letters to each of the rows and to each of the columns. The letters are then grouped in such a fashion that the combined averages of usage for each row or column are distributed as evenly as conditions permit.

While squares of varying sizes may be thus constructed, attention was given to the ultimate intended use of the system, namely, the recording of the contents of a square in a single row of an 80-column IBM card. The largest square, therefore, that could be accommodated is an 8 by 8 square, requiring 64 positions across the card, thus leaving 16 columns for recording serial numbers and other information.

While the spelling of conventional words would be possible with the new method, it was also essential that it be equally adaptable to more compact and less redundant schemes of spelling. It should also be possible to encode combinations of numerals such as serial numbers and numeric codes.

Considering first alphabetic schemes, attention was given to the following spelling schemes:

1. Conventional Spelling
2. Consonant Code
3. Significant Letter Code
4. ELCO Code
5. Self-Demarcating Word Code

These schemes will be described briefly before discussing the procedure of arriving at randomizing squares.

Special Consonant Code. The following code is proposed to provide a simple systematic method for deriving code words from the original words. It is a variation of the conventional "Consonant Code" which normalizes words by deleting all vowels, the letters W, H, Y and the duplicate in double letters. However, in order to keep words from being consumed by this process, as in the case of the word "WAY," the following procedure is proposed:

"Starting from the right, strike out all duplicates of double letters and the U of QU. Then starting from the right again strike out vowels and the letters W, H, Y

but stop short if this process reaches a fixed remainder of 3 letters, for example. (OLIVE is reduced to 'OLV', WAY remains 'WAY')."

The consonant code is particularly useful for encoding proper names in that it overcomes many of the variations to which such names are subjected. Additional conventions may be introduced to handle the usage of K, CK, C, of TS, TZ, Z and other common variations.

Significant Letter Code. This encoding scheme is a process of abbreviating common words or names to a fixed minimum of the given letters.³ The theory is that less frequently used letters provide greater differentiation among abbreviations. Words are therefore systematically reduced by eliminating letters in accordance with a letter use frequency table. Of the letters in a word, the one having the highest frequency ranking is dropped first, the one next in frequency is then dropped, and so on, until the fixed minimum of remaining letters has been reached. Double letters are treated as single letters and the U of QU is disregarded. Among similar letters the last one is dropped first.

A maximum of 4 letters appears to be sufficient for the average application. The first letter of a word is retained as being significant because of its position and is therefore excluded from the reduction process. This may aid in identifying and indexing the abbreviations.

To a degree, this process results in a more even usage of the letters of the alphabet and therefore promotes random distribution of superimposed marks.

A frequency scale which might be used for the reduction process is that compiled by R. T. Griffith as published in the *Journal of the Franklin Institute*, as follows:

E	T	A	O	N	I	S	R	H	L	D	C	U	M	F	Y	W	G	P	K	B	V	X	J	Q	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Examples:

APpaRatUs = APRU SeQueNCe = SQNC

"ELCO" (Eliminate and Count) Code Words. This code is proposed to furnish an added degree of differentiation over the Significant Letter Spelling method. The procedure for deriving an ELCO code word is as follows and is applied only to words of more than 4 letters:

"Starting from the right strike out all duplicates of original double letters and the U of QU. In doing so write over each of the letters thus eliminated the number assigned to it in the letter frequency scale. If more than 3 letters remain, strike out

³ Carl A. Cline, "An Edge-Notched Index Card System for Mechanical Sorting", Paper presented to the Div. of Chem. Literature, American Chemical Society, New York, September, 1954.

additional letters, with the exception of the word-starting letter, in descending order of frequency ranking as given by the letter scale. Again write scale number over each of the letters stricken out. Of similar letters, strike out the one farthest to the right. When the word has thus been reduced to 3 letters, add up the values of the letters eliminated. If the total is over 26, deduct 25 as often as necessary. On the frequency scale find the letter which corresponds to the value of the derived total. This letter becomes the 4th letter of the ELCO word, the first letter being the starting letter of the original word and the 2nd and 3rd letters being the most significant of the remaining letters of the word. In case the original word has 3 or less letters, add the letter Z, standing for 'zero', on the right to bring the word to 4 letters."

The following examples have been derived with the aid of the Franklin Institute scale previously given and numbered for the computation of ELCO code words:

ELCO		Significant Letter Spelling (for purposes of comparison)
35 1	= 9 = H	
CanCeR	= CCRH	CNCR
CONCERN	= CCRF	CNCR
CONCERT	= CCRC	CNCR
CONCRETE	= CCRU	CNCR
PATENT	= PANN	PATN
PATENTEE	= PANS	PATN
FAT	= FATZ	

Self-Demarcating Code Words. Such code words are made up of sequences of consonants and vowels in such a manner that several of the code words can be written side by side without the need of separating the words by special marks. All of these code words begin and end with certain of the consonants. Three-letter words have a vowel in the middle while four-letter words have either two vowels or one vowel and the letter L or R. There are some exceptions to these rules.⁴ As will become apparent later, this systematic spelling of code words is particularly adaptable to the new system and contributes to the reduction of spurious matches.

Design of an 8 x 8 Randomizing Square for Letters Only

The aforementioned methods of coding have been made the basis for the manner in which the letters of the alphabet have been grouped in the indexes of rows and columns of the encoding matrix. The problem, therefore, was to come up with an arrangement that would achieve a comparable degree of random distribution for all of the five modes of spelling.

Since in an 8 by 8 square at least three letters have to be assigned to each row or column, precautions had to be taken to insure differentiation

⁴ H. P. Luhn, "Self Demarcating Code Words," April (1953), IBM Research Center, Yorktown Heights, N. Y.

of spelling among the three letters of each group. Obviously, if the same groupings were applied to the vertical as well as to the horizontal index, the letters of a group would be treated in identical fashion. Therefore, a given chain would represent all of the words that could be derived by permuting the various letters of all the groups in the chain. This would also mean that the sequence of the letters of a pair would not be expressed since, for example, the pair AB and the pair BA would be represented by the same intersection. This situation was substantially overcome by grouping the letters in the vertical set differently from those in the horizontal set so that no two or more letters of a group in a horizontal set would re-occur in a group in the vertical set. This however required the creation of two sets of groups and the optimization of randomness in each set.

As far as statistical information is concerned on the frequency of usage of letters in the English language, the table reported in the *Journal of the Franklin Institute* was used. However, since this table did not differentiate between the various positions of letters within a word, it was deemed advisable to modify its values to reflect the frequency of starting letters on the basis of listings in *Webster's Collegiate Dictionary*.^{*} An average of four letters per word was assumed for the usage of the matrix and a new table of values was computed on the basis of one starting letter and three average letters. In Table 23-1 the three sets of values are shown side by side whereas Table 23-2 lists the newly derived values in descending order.

The grouping of the letters into the final arrangement was made to produce, as nearly as possible, a reasonably even distribution of combined averages of letters:

1. For fully spelled-out words.
2. For the Consonant Code.
3. For the Significant Letter and the Elco Code.
4. For the Self-Demarcating Code Words.

The following rules were observed in distributing the letters:

1. To try for an optimum distribution of consonants for the Consonant Code.
2. To assign the vowels and the letters L and R singly, that is, one to a row or column, in order to optimize the distribution of the inner letters of the Self-Demarcating Code words.
3. To pair each of the letters L and R with one of the consonants W, H, Y.

The letter groupings arrived at on the basis of the above considerations is given in Table 23-3. The left-half of this table shows the distribution

^{*} A more effective means of deriving frequency tables would be by way of analysis of word usage in the particular field a system is to serve.

LETTER FREQUENCY COMPUTATION

TABLE 23-1

TABLE 23-2

Frequency			Combined Average $\frac{3F + 1W}{4}$	Listing in the New Order of Frequency
Letter	Franklin	Webster		
A	8.1	6.4	7.5	E 9.5
B	1.5	5.4	2.5	T 8.5
C	2.9	9.7	4.5	A 7.5
D	3.7	5.1	4.0	S 7.5
E	12.1	3.7	9.5	I 6.5
F	2.3	4.3	3.0	O 6.0
G	2.0	3.3	2.5	N 6.0
H	5.4	3.9	5.0	R 5.5
I	7.3	3.9	6.5	H 5.0
J	0.1	0.9	0.5	C 4.5
K	1.7	0.8	1.5	D 4.0
L	4.0	3.4	4.0	L 4.0
M	2.5	5.2	3.0	P 3.5
N	7.3	2.0	6.0	M 3.0
O	7.5	2.3	6.0	F 3.0
P	1.9	8.6	3.5	B 2.5
Q	0.1	0.6	0.2	U 2.5
R	6.0	4.7	5.5	W 2.5
S	6.1	12.0	7.5	G 2.5
T	9.3	6.0	8.5	Y 2.0
U	2.8	1.6	2.5	K 1.5
V	1.0	2.1	1.5	V 1.5
W	2.1	3.2	2.5	J 0.5
X	0.1	0.1	0.1	Q 0.2
Y	2.1	0.3	2.0	Z 0.2
Z	0.1	0.3	0.2	X 0.1

for one axis and the right-half shows the distribution for the other axis. Columns b, c, and d contain the consonants except WHY. Column b contains the first eight consonants of Table 23-2 listed downward in descending order, except for S and R. Column c contains the next eight consonants of Table 23-2 in descending order listed upward. In column c' of the right-hand portion of the table the sequence of the letters has been changed by transposing the entries in column a. Thus, 1 and 2 have been interchanged, as well as 3-4, 5-6, and 7-8. The vowels and WHY were then entered into column a in accordance with the above rules. The letters W and Y in the left-hand side of the table have been paired with R and L, respectively. Column a' of the right-hand side was then derived from that of the left-hand side by transposing the first four and the last four entries of column a on the left. In column d and d' were then entered the two

TABLE 23-3. LETTER GROUPING FOR AN 8 x 8 RANDOMIZING SQUARE.

	a	b	c	d	Cons. Code b/c/d	Total	a'	b'	c'	d'	Cons. Code b'/c'/d'	Total
1	U 2.5	T 8.5	Q .2		8.7	11.2	O 6	T 8.5	J .5		9	15
2	I 6.5	S 7.5	J .5	X .1	8.1	14.6	A 7.5	S 7.5	Q .2	Z .2	7.9	15.4
3	W 2.5	R 5.5	V 1.5		7	9.5	Y 2	R 5.5	K 1.5		7	9
4	E 9.5	N 6	K 1.5		7.5	17	H 5	N 6	V 1.5		7.5	12.5
5	O 6	C 4.5	G 2.5	Z .2	7.2	13.2	U 2.5	C 4.5	B 2.5		7	9.5
6	A 7.5	D 4	B 2.5		6.5	14	I 6.5	D 4	G 2.5		6.5	13
7	Y 2	L 4	F 3		7	9	W 2.5	L 4	M 3		7	9.5
8	H 5	P 3.5	M 3		6.5	11.5	E 9.5	P 3.5	F 3	X .1	6.6	16.1

consonants X and Z, adding them to combinations which contain a vowel. They were not combined with U, however, because of the degree of significance this vowel has in the Significant Letter and Elco Code. On the right-hand side of the table the association of these two letters was varied so that they would not appear together with the letters they were combined with on the left-hand side.

The combined averages are given alongside the columns in Table 23-3, first for the consonants of the consonant code and then for the whole group. It is apparent that the grouping favors the consonant combination for the consonant code, the combined averages ranging from 6.5 to 9. As far as self-demarcating code words are concerned, a reasonably even distribution is assured by having each group contain outside letters and inside letters in the same proportion. The total averages range from 9 to 17.

While it is felt that the foregoing arrangement is a reasonable solution for the application it was designed for, other groupings may be in order when dealing with different codes or languages. In this connection reference is made to "Interlingua" because of its application to scientific information. In any case the above-mentioned procedures will facilitate the construction of an optimum matrix.

In order to arrive at the final recording square, the two lists represented in Table 23-2 have been arranged at right angles to each other to designate intersecting rows and columns. The left-hand list has been used as indexes for the rows and the right-hand list as indexes for the columns. The resulting arrangement is shown in Figure 23-1. For convenience of reference the letters within a group and the groups have been rearranged in alphabetical sequence, since the actual position of a group with respect to other groups is immaterial.

For example, an entry has been made in the square of Figure 23-1 of

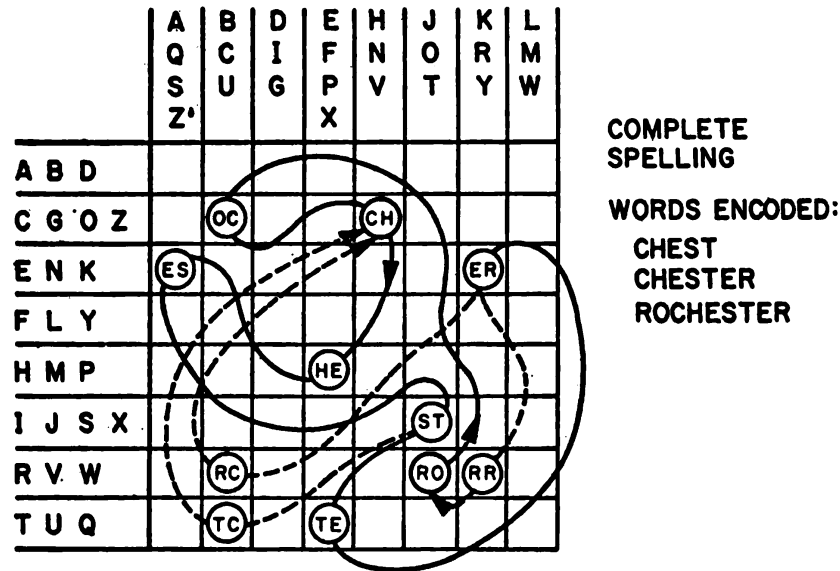


Figure 23-1.

the words "CHEST," "CHESTER," and "ROCHESTER" with all the letters spelled out. The various marks have been interconnected to indicate the chain formed by the sequence of the pairs of letters. The around-the-end portion of the spelling in each of the three cases has been indicated by dotted lines and the affected marks have been labelled accordingly. The arrows in each case point to the beginning of each of the words. The function of end-around spelling to differentiate the three words becomes apparent in this diagram.

An example is also given of each of the patterns created by consonant spelling, significant letter spelling, and self-demarcating code spelling, as applied to an abstract of the same subject matter given as a single sentence in Table 23-4. The sentence has been written in eight lines, each of which contains a "notion" which might have been chosen by the editor as a differentiating element. The example is given merely to illustrate the various word codes, and not a particular method of abstracting.

TABLE 23-4

Abstract	Notion	Cons. Code	Sign. Letter Code	ELCO Code	Self Dem. Code
The recording of information by superimposed code combinations and the selection of desired documents by a statistical method of scanning and matching	write inform merge code select book approximate scan	WRT NFRM MRG COD SLCT BOK PRXMT SCN	WRIT IFRM MERG CODE SLCT BOOK APXM SCAN	WRIA IFMW MRGT CODE SLCO BOOK APXS SCAN	WRIT NIF MERJ KOD SLEX BOOK PROX XAN

The column to the right of the complete sentence gives the notional terms assigned by the editor; the word codes for these are listed in a separate column for each of the four spelling methods.

The pattern derived by consonant spelling is shown in Figure 23-2. The various words have been numbered and the marks in the square have been identified by these numbers to facilitate tracing the procedure of chain spelling. The spelling of the 28 letters resulted in 25 marks, 3 of which are double entries.

The pattern created by significant letter spelling is shown in the lower left of Figure 23-2. In this case the 32 letters resulted in 25 marks of which 5 are double entries and 1 a triple entry.

The pattern of self-demarcating code spelling is shown in the lower right of Figure 23-2 where 29 letters resulted in 24 marks including 5 double entries.

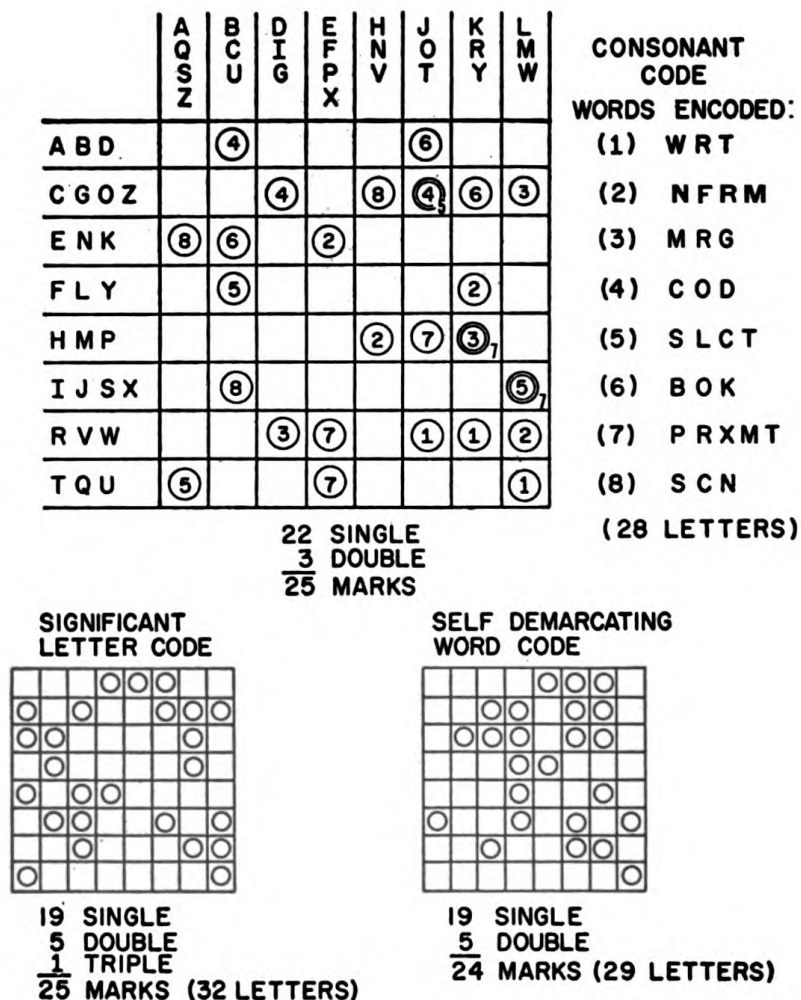


Figure 23-2.

Design of a 10 x 6 Randomizing "Square" for Mixed Symbols

The above design procedures were directed at the creation of squares with alphabetic indexes. It is apparent that any other set of symbols could be adapted for randomizing and for the distribution of code marks by the method of chain spelling. The use of numerals for serial numbers and numeric codes makes it desirable to enter such information in superimposed fashion, preferably in conjunction with alphabetic information.

A scheme which permits entries in mixed symbols will be described. For the purpose of this example it has been assumed that the frequency of usage is the same for all numerals. It has also been assumed that the size of the square should be such that the code patterns could be recorded on a punched card very much like the previously developed square and to a similar extent.

Because there are 10 numerals, the format of this square has been chosen to provide a balanced distribution of the 10 entries in the indexes of the rows as well as of the columns. A 10 x 5 square accomplishes this by assigning one numeral each to the 10 rows and the 5 pairs 05, 16, 27, 38, 49 to the five columns. In order to identify the end of a chain, an "End Mark" column is provided for entering the last character of a term instead of pairing it with the first character of a term as in the previous examples. This 6th column brings the size of the square to 10 fields high and 6 fields wide.

The letters of the alphabet have been grouped into two index sets, one for the 10 rows and one for the 5 columns. The same principles of distribution have been used as in the previous examples. The letters "O" and "I" have been arranged to coincide with the numerals "0" and "1" to overcome confusion between these symbols.

The groupings of characters and the combined averages of letter usage are given below.

Row Index				Column Index			
		Average All Letters	Cons. Code			Average All Letters	Cons. Code
COZ	0	10.7	4.7	FHOQS	05	21.7	10.7
DIJ	1	11	4.5	GIKLP	16	18	11.5
AMV	2	12	4.5	ACTWX	27	23.1	13.1
BL	3	6.5	6.5	BDRUV	38	16	13.5
EFK	4	14	4.5	EJMNYZ	49	21.2	9.7
GR	5	8	8				
HT	6	13.5	8.5				
NW	7	8.5	6				
PQU	8	6.2	3.7				
SXY	9	9.6	7.6				

		F H O Q S	G I K L P	A C T W X	B D R U V	E J M N Y Z	END MARK
		0 5	1 6	2 7	3 8	4 9	
C O Z	0	(2)	(2)				
D I J	1		(3)	(1)			
A M V	2			(1)	(4)	(3)	(1)
B L	3			(3)		(1)	(2)
E F K	4	(1)	(1)				
G R	5	(2)				(5)	
H T	6			(1)	(2)		
N W	7					(1)	(5)
P Q U	8		(4)	(2)			
S X Y	9			(1)			(3)

18 SINGLE
 4 DOUBLE
 1 TRIPLE
 23 MARKS

NUMBER CODES
 (1) 34F1746X2
 (2) PC5063
 (3) 2,416,379
 (4) 828L
 (5) 55N
 (30 CHARACTERS)

Figure 23-3.

Figure 23-3 shows the pattern created by an example of 5 number code entries comprising 30 characters.

For purposes of comparison the word code entries of the 8 x 8 square have also been entered in the 10 x 6 version as shown in Figure 23-4.

Checking Schemes

Numeric codes which utilize self-checking features for eliminating transcription errors⁵ are particularly effective in superimposed coding schemes. This is so because the check digit, commonly used in such schemes, is systematically computed and consequently acts as a unique differentiating element. It will therefore be well to take advantage of this device as it tends to reduce the rate of spurious matches. These checking systems are equally adaptable to mixed and purely alphabetic codes.

⁵ "Self-Checking Number System", International Business Machines Corporation, New York City, Publication: Form No. 22-6022-0. See also U. S. Patent No. 2,731,196 to H. P. Luhn dated Jan. 17, 1956.

		F H O Q S	G I K L P	A C T W X	B D R U V	E J M N Y Z	END MARK	
		0 5	1 6	2 7	3 8	4 9		CONSONANT CODE
C O Z	0	(4)	(6)	(5)	(4)	(8)		(1) WRT
D I J	1						(4)	(2) NFRM
A M V	2			(7)	(3)		(2)	(3) MRG
B L	3	(6)		(5)				(4) COD
E F K	4				(2)		(6)	(5) SLCT
G R	5		(3)	(1)		(2)	(3)	(6) BOK
H T	6						(7)	(7) PRXMT
N W	7	(2)			(1)		(8)	(8) SCN
P Q U	8				(7)			
S X Y	9		(5)	(8)		(7)		(28 LETTERS)

23 SINGLE
1 DOUBLE
1 TRIPLE
25 MARKS

SIG LETTER CODE

○	○	○	○		
○		○		○	
	○			○	○
○		○			○
			○		○
	○			○	○
			○		○
		○			○
	○	○		○	

16 SINGLE
8 DOUBLE
24 MARKS (32 LETTERS)

S.D.C.

○	○	○	○		
○		○			○
				○	
○				○	
○	○	○		○	
○	○			○	
			○		○
	○	○		○	
		○			○
	○	○		○	

21 SINGLE
4 DOUBLE
25 MARKS (29 LETTERS)

Figure 23-4.

Single-Row Recording on Punched Cards

It was considered more convenient to demonstrate the new system and the various examples first in their original two-dimensional form. As mentioned earlier, one of the objectives was the creation of single-row recordings of the patterns contained in the squares, each row representing a "sentence." This is achieved by laying the rows of the squares end-to-end in a given order as illustrated in Figure 23-5 which shows an 80-column IBM card for the combination alphabetic and numeric scheme. The 60

If the code words used in the preceding example were to be spelled out, at least 40 characters and division marks would be required. When using self-demarcating code words, this may be reduced to 30. If a 6-bit code were used to record this, the space required would be the equivalent of 30×6 or 180 bits. The recording of a 10×6 randomizing square, on the other hand, would require a space for only 60 bits. This is the space needed for 10 characters so that in this particular instance a reduction of 4 or 3 to 1 could be realized. Because of this reduction, less storage space and functional capacity would be required and processing time would be materially shortened. Whether these reductions would pay off depends, of course, on the relative merits of the statistical, and the discrete searching methods as applied to a given situation.

If it is desired to record randomizing squares on 7-channel punched or magnetic tape, this could readily be done by sectioning the squares into 6-bit strips and by recording these across the tape the same way as normal characters are recorded. If desired, a bit count can be made for each row and a redundancy bit can be added in the 7th channel where required. In the case of the 10×6 square, no rearrangement of the pattern is required.

Conclusion

The principle of distributing code entries by the use of randomizing squares and the principle of chain spelling have been demonstrated with the aid of letter frequency tables derived from common English literature. In order to obtain greatest efficiency, it will be advisable to compile special frequency tables for scientific literature.

The principles described here may equally well be applied to any foreign language and artificial languages such as "Interlingua," provided "letter use frequency tables" are available for these languages.

Superimposed coding offers many advantages where recording space is at a premium. For this reason it is being used extensively in marginally punched card systems. For the users of such systems the randomizing square method of recording may offer additional advantages.

Chapter 24

INDEXING AND INDEX SEARCHING

E. J. CRANE AND CHARLES L. BERNIER

Chemical Abstracts Service

The Ohio State University, Columbus, Ohio

Introduction

The inclusion of a chapter on indexing and index searching in a book on punched cards seems natural since it is coming to be recognized that punched-card and similar mechanized searching systems for finding information in documents can be considered as manipulative indexes. These indexes must be manipulated by correlation of terms, numbers, punched holes, etc., in order to select the relevant information or, more usually, references to it. Published subject indexes, by way of contrast, can be considered as nonmanipulative, or manipulative only to the slight extent represented by the turning of pages and by reading. Both types of indexes, manipulative and nonmanipulative, are built on many of the same principles, which involve: selection of terms to represent the documents indexed, vocabulary control, and the like.

Manipulative indexes are an important, new development. They obtain their selectivity of documents or references by correlation of two or more terms taken simultaneously to generate, in effect, more specific subjects and classes of subjects. That is, the user of these indexes greatly limits the number of documents, etc., which he must examine by using, in correlation, any combination or permutation of terms which he chooses from the vocabulary to the index.

It has been thought that the selectivity achieved by correlation of terms could be obtained only through manipulation at the time of use. The very large number of permutations and combinations of vocabulary terms has seemed to preclude the recording of them in a nonmanipulative form of economically, or even physically, practical size. It now seems possible to produce nonmanipulative correlative indexes in book form. This can probably be done by limiting the recorded combinations of vocabulary terms to those leading to actual (and not potential) documents, by limiting the number of permutations by alphabetizing, by controlling the scattering from partial combinations by various syndetic devices, and by limiting the vocabulary by the use of systematic nomenclature, rhetorical tropes, thesauri, syndetic devices, or some combination of these. Such alphabetical correlative indexes in book form would seem to have many advantages over manipulative indexes. With such alphabetical correlative indexes it

now appears possible to provide, in book form, the results of all significant searches made in many types of mechanized documentation systems and to perform most of the functions of manipulative correlative indexes more efficiently. From the viewpoint of this paragraph a chapter on indexing and index-searching seems even more fitting.

Manipulative or mechanized systems (proposed or actual) have brought with them a number of serious problems, all of which, to date, have not been completely solved for correlative indexing of large documentation systems dealing with broad fields of knowledge to be used by others than the creators of the systems. Among these problems are:

- (1) The imperceptible loss of relevant information caused by correlation of too many terms simultaneously.
- (2) Blank sorts resulting from searching for nonexistent classes.
- (3) The unavoidable selection of unwanted, irrelevant information.
- (4) Confusion of meaning because relations among the vocabulary terms selected were difficult if not impossible to show completely, i.e., because many of the systems lacked morphemes.
- (5) Deficiency in effective "browsability" and immediate suggestion of related and substitute information.
- (6) Necessity for manipulating the system before relevant documents, etc., could be located. I.e., the results of all significant correlations were not immediately available without manipulation.
- (7) The relative bulkiness of the recording media and associated apparatus when compared with indexes in book form.
- (8) The probable relative costliness.
- (9) Delays caused by the necessity of manipulating the system or of communicating with a "documentation center" in order to get the answer to a question.
- (10) The economic restrictions imposed by the more costly systems, which restrictions reduce the total amount of information communicable.
- (11) The bringing of pertinent vocabularies of system and searcher into coincidence.
- (12) The facilitation of generic searches.

While this chapter will not provide tested solutions to these problems, most of which are largely associated with manipulative indexes, it will give tested methods of subject indexing which can be used until adequate solutions have been discovered.

Manipulative indexes using punched cards have proved successful for narrow fields of knowledge, especially when the sole user is the indexer. In small systems of this type there is no longer the problem of bringing vocabularies of system and searcher into coincidence because the system was indexed by the only searcher. The small size and probable highly spe-

cialized nature of the collection minimize the problems relating to cost, delays, irrelevant information, bulkiness, generic searches, and the necessity of manipulation. The problems of browsability, loss of information, blank sorts, and confusion of meaning are greatly reduced because the owner (and builder) of the system, after the considerable effort of acquiring and indexing it, already knows fairly closely what it contains.

In general, it is true that machines have done much to reduce manual labor and improve products. Machines are also helpful in various ways in reducing mental effort, as by the use of calculating devices. Since the early forties there has been growing interest in the possibility of utilizing machines in the recording and searching of scientific information. The principal methods tried have involved punched cards. Electronic devices have also entered the investigational picture.

Of course, machines have long aided in the recording of information. Typewriters and printing presses are machines. *Chemical Abstracts* is indexed by the use of magnetic recorders, plus transcription of the magnetic record on electric typewriters. Well informed and highly trained indexers dictate index entries to save their time and produce a more legible product. This procedure separates technical and clerical work.

An object of high enthusiasm in the field of literature mechanization has been the hope that literature searching could be so facilitated that it would be a matter of pressing buttons and the like to obtain needed references or even the information itself.

No comprehensive mechanization system has been found as yet for a field as broad as the whole of chemistry and the related sciences in which chemistry is often used. The most important American project of this sort has been that of the Chemical-Biological Coördination Center sponsored by the National Research Council, National Academy of Sciences in Washington, D. C. It is limited to chemical compounds (about 60,000 of them) for which certain recorded biological properties have been correlated with coded structural elements. Dealing with chemical compounds by means of punched cards is more promising than dealing with less definite information (chemical phenomena, for example) by such means.

The effort for greater use of machines in dealing with the scientific literature, such as machines aiding in the production of indexes, is worthy and should be continued. The objective should be kept high and the limitations should be recognized. Machines cannot think, but they can tirelessly do many things with great rapidity and accuracy. Machines cannot provide nonnumerical information which human beings have not, with forethought, put into them.

For many applications, it may be more expensive to record, distribute, and use information on cards than to perform the same operations by means of indexed journals. Because of the expense factor, the latter method

will probably continue in use even if mechanized literature becomes completely successful. Often individuals will not be in a position to acquire mechanized information systems.

There are many more or less specific purposes in the field of scientific information for which machines can provide effective help. Our purpose in emphasizing limitations is to suggest the wisdom of substituting an element of precautionary realism for some of the enthusiasm which is leading to higher claims for literature mechanization than present attainment has provided. The enthusiasm should continue for the objective is most worthy.

The ideal literature storing and searching system (manual or mechanized, manipulative or nonmanipulative) would prevent the loss or concealment of hard-won facts and it would easily assemble all pertinent facts in answer to a question. If the searcher's question were not exactly the right one, the ideal system would help him by drawing on its own resources, for example, of synonyms or related words. Such a system might help the searcher by giving to him abstracts or articles rather than references to these. The ideal system would give rapid answers to generic questions such as, "What biological properties of olefins were studied last year?"

In the discussion to follow an effort will be made to give the characteristics of a good subject index, to tell how such an index can be built, to bring in from time to time relations to mechanized systems, and to describe an effective procedure for index searching. Both of the authors of this chapter have gained most of their indexing experience in the office of *Chemical Abstracts* and ask the privilege of referring rather frequently to the indexes and the indexing practices of this service.

Indexing

An index is a pointer or key which directs the searcher to recorded information. A good subject index is also a kind of inanimate memory. The human memory and well-indexed compilations are similar in that both store information in such a way that it can be recovered upon demand. Indexes bring together like information and, with their cross references, help in the correlation of data. The whole picture of what is happening in the field covered by an abstract journal can be gained in outline form from its indexes.

There are, of course, various kinds of indexes. *Chemical Abstracts* publishes indexes devoted to (1) authors, (2) subjects, (3) chemical compounds listed by formulas, (4) patent numbers, and (5) organic ring structures. This discussion will be limited to subject indexes and subject indexing. The subject index is the one used most in chemical and other fields. The subject part of an important chemical index is always the first to show signs of wear in libraries.

What is a good subject index? It is one that will serve as a reliable means

for locating, with a minimum of effort, every bit of information in the source covered which, according to the indexing basis, that source contains. To meet this test an index must be accurate, complete, sufficiently precise in the information supplied, and so planned and arranged as to be convenient to use. Existing indexes fall far short of this ideal in many cases and, of course, somewhat short of it in all cases.

Completeness, usually attained in author and patent-number indexes, can only be approximated in subject indexes, and is frequently far from being reached. For most kinds of publications thorough indexing is highly important, but even subject indexing can be overdone. An index may be reasonably complete from one point of view and not from another. For example, a chemist may have occasion to make use of a publication on bacteriology, the index to which is adequate from the point of view of bacteriology but incomplete from the point of view of chemistry. That is reasonable to expect; it should be kept in mind in making searches. Completeness can be considered both with reference to headings and with reference to modifications (page 518).

Some indexes to periodicals, particularly word indexes, are merely indexes of titles of papers or of abstracts, as the case may be. These are always incomplete. Titles frequently do not tell, even in very general terms, the whole story as to the contents of papers. Very often papers will contain new data with reference to specific substances not mentioned in the titles (possibly referred to in a general way, as by "some hydrocarbons"); these data should certainly be made available by means of specific index entries. Furthermore, many papers, particularly in conclusions, contain significant information with reference to more or less abstract subjects which are not brought out in the titles at all, as the relation between color and chemical constitution.

Of course, the need for full indexing varies somewhat with the nature of the publication covered. It is particularly important for an abstract journal to be thoroughly and properly indexed. Abstracts should be prepared from the indexing point of view. In other words, they should contain all of the information in papers covered which should be indexed, and the index entries should then be made. As an indication of how the abstractor and indexer, working in coöperation, endeavor to make the record complete, the first part of rule 32 is reproduced from "Directions for Abstractors and Section Editors of *Chemical Abstracts*":

"32. Since *Chemical Abstracts* is intended to be a complete and permanent record of all chemical work, it is very important that abstracts should contain or make specific reference to all of the information in articles that is suitable for index entries. This would include every measurement, observation, method, apparatus, suggestion, and theory that is presented

as new and of value in itself. All new compounds and all elements, compounds, and other substances for which new data are given should be entered in abstracts. . . ". The question of precision in indexing will be discussed below in dealing with modification writing.

Convenience in use has not always been considered in the building of subject indexes. It is important since literature searching is a time-consuming task even under the best of conditions. The movement in the direction of mechanization has convenience in use and time saving as principal objectives. It is possible by efficient modification writing, systematic arrangement of entries, and judicious selection of printing form and style greatly to facilitate the use of a subject index. The general quality of the indexing and the use of cross references, which will be discussed later, are also important factors. Probably the most effective printing form is the so-called "entry-a-line" form with alphabetic arrangement of modifications the significant word of which has been brought to the front¹.

The greatest weakness in scientific literature lies in the existence at many points of inadequate and poorly constructed subject indexes. The importance of such indexes has not always been realized. However, the trouble probably lies in the lack of realization that only a trained and experienced indexer can be expected to be able to make a good subject index. Indexing is a science and an art in itself. Subject indexing has often been attempted by individuals whose only qualification was a knowledge of the field to be covered. Not even authors are qualified to index their own work unless they are equipped for the task by training and experience. To become a satisfactory indexer for a given publication one must have certain general qualifications for the work which can be acquired by experience, in addition to a considerable acquaintance with the whole branch of knowledge involved, ability to comprehend fully the contents of the publication, and familiarity with the principles and practices of indexing. Good taste, good judgment, conciseness, and liberal and comprehensive thought are also necessary. Above all, one needs what may be called the "indexing sense," that is, the ability to feel instinctively, at first glance, what and how subjects should be indexed in all their ramifications².

The indexer uses words as tools for the location of subjects. Many so-called subject indexes are really indexes of words instead of subjects. There is a vast difference. Word indexing leads to omissions, scattering, and unnecessary entries. After the most suitable word or group of words from the indexing point of view has been chosen for a heading or even, at times,

¹ A good discussion of the various forms of subject indexes, with examples, as well as of indexing in general is to be found in *Bull. No. 779*, Albany, University of State of New York (May 1, 1923).

² Nicholas, J. B., *Library J.*, 17, 406 (1892).

for a modifying phrase, this should be used consistently no matter what the wording of the text may be.

Since much of chemistry has to do with chemical compounds, many of the entries in a chemical index are the names of these compounds. It is clear at once that the use of good systematic nomenclature is important in subject indexing in the chemical field, as is indeed true in recording information in any field. Nomenclature is a special problem in chemistry because of the enormous number of known chemical compounds, the properties, reactions, and uses of which are likely to be studied frequently, and because of the fact that many thousands of new compounds are prepared and studied each year. The subject of chemical nomenclature will not be considered here, but sources of helpful modern information on this subject are listed at the end of the chapter.

To function efficiently as a key or guide or inanimate memory, an index must have entries (the units from which an index is built) which are easy for the user to find and understand. This seemingly simple and obvious requirement of all good indexes is the one that causes much of the trouble in actual index building. Entries may be difficult to find not only because they may be placed at some unexpected position in the index (as occurs frequently in classified indexes), but also because they may be expressed by unexpected or unfamiliar words.

In building indexes the placing of entries at unexpected positions can usually be avoided by restricting classification to a minimum, or by avoiding it altogether. What is logical to one person may be illogical or unexpected to another. This is a very important point and deserves an illustration.

Where, for example, should one look for vitamin C in a *classified* index? Would this compound be found under: Chemical compounds, Acids, Reducing substances, Biological materials, Food, Dietary factors, Nutritional accessories, Vitamins, Growth substances, Sugar derivatives, Enediols, or Lactones? Vitamin C would logically fit into all of these classes. The user of a classified index might find vitamin C by looking under all possible classes. This method, however, would be inefficient and it might be futile because any material thing can usually be placed easily and logically in many more classes than the above sample list indicates. A better way of locating vitamin C in a classified index would be to consult the classification scheme for the index and to select from this scheme the more probable places to look. The most efficient way would be to consult an alphabetical index to the classified index and discover exactly where the indexer put vitamin C. The use of a classified index with a supplementary alphabetical index is in most cases probably not so efficient as the direct use of an alphabetical index to locate the information desired. Hence, the conclusion has been reached that classification should usually be avoided.

The use of unexpected or unfamiliar words in an alphabetical index is inadvisable. The good indexer places entries where the index user is likely to look first. This means that the most common synonym should be selected even though one might be tempted to choose another. It might be argued that vitamin C is not a good chemical name since it gives no clue as to the chemical nature of the compound (it is not an amine) and is not uniquely essential to life (not necessary for some forms of life at least). Ascorbic acid is a much better chemical name. However, if most people look under *Vitamin C*, that is a better place for index entries than under *Ascorbic acid*. An indexer must constantly keep in mind that he is not constructing a logical edifice, but rather a device (a key) which will lead the user to the information he desires as quickly and as surely as possible.

It also happens sometimes that it may not be strictly accurate by laws of logic to group certain entries, yet good indexing demands it. A case in point is the entry of protium studies under *Hydrogen*. Strictly speaking, hydrogen is a mixture not only of protium, deuterium, and tritium, but also of ortho- and para-hydrogen and (in a commercial sense) other gases as well. Yet good indexing seems to demand that protium studies be entered under *Hydrogen* because most people will look there for information on the manufacture, reactions, properties, and uses of protium in a relatively pure state. If it were possible for the indexer to separate all of the studies on protium from those on hydrogen (the mixture) and to index these separated entries at *Protium*, then the index user would be forced to look under two headings for related or identical information since the reactions, properties, uses, etc., of protium and of hydrogen are usually identical. Such indexing would not improve efficiency in using the index.

Cross References

A part of the difficulty in the use of indexes comes from the inability to locate the term under which the information is placed, as explained above. This situation can be greatly helped by the use of cross references, common practice in indexing. A cross reference is an index entry of a special type that directs the index user to turn to another part, or suggests that relevant information will be found in other locations.

Synonyms are handy words for poets, but are a hindrance to indexers and index users. Cross references are of special help in handling the problem created by synonyms and in correlating the various related entries. When two synonyms, such as thiamine and vitamin B₁, are about equally used, then a cross reference, as *Thiamine*. See *Vitamin B₁*, is needed. Also, two concepts can be combined and indexed at one place more easily by the help of cross references, even though the two concepts are not exactly synonymous. Thus, *pH* can be combined with *Hydrogen-ion concentration*.

Correlating cross references are valuable in introducing the index user

to closely related information. *Aluminum alloys*. (See also *Duralumin*) is an example.

The use of certain cross references can save index space. The cross reference, *Furnaces*, for glass manuf. or melting—see *Glass*, will save the space taken by entries at the heading *Furnaces* each time a glass furnace is indexed. Experience has proved that this type of cross reference should be used with caution since important correlations may be lost or made more difficult to locate. This type is especially valuable in handling the indexing of subjects bordering on the main subject of the index, such as mathematics and physics, which border on chemistry, when great detail and heavy indexing are not desired for the borderline subjects.

Inverted Cross References

Once a cross reference has been made, it must operate consistently. In indexes to the current literature which come out periodically the consistent use of cross references presents a problem, especially if the index is of considerable size. In an expanding cross-reference system it has been found that cross references occasionally get in which operate at cross-purposes. This happens more often when several indexers contribute cross references to the system. In order to prevent inconsistencies, and to ensure consistent operation of the cross-reference system, several methods of cross-reference control can be used. One of the more effective is the inverted cross reference. For *Iron*. (See also *Steel*), the inverted cross reference would be *Steel, Iron*, (See also ---). This inverted cross reference is typed onto a card and alphabetized at *Steel* so that the editor at that heading can discover whether it operates at cross-purposes with any cross references going out from the *Steel* heading, and decide whether the cross reference represented by this inverted cross reference is to be allowed to operate at a given time. The editor can record his decision most conveniently by a mark on the inverted cross reference.

Index Structure

A word about index structure is due at this point. The word or phrase selected to act as guide to the subject, concept, author name, etc., is called a *heading*, and is helpfully printed in bold-face type for emphasis; correct indentation of the rest of the material in the printed index can also be used to help set off the heading.

The *modification* is anything more that is said about the heading in an individual entry. Modifications are usually used in indexes which lead directly to the literature; they may not be required by indexes to mechanized systems since the system may be used in some measure to perform

the work done by the modification. In the entry, *Cyclohexane*, chlorination of, 3148*i*, the phrase "chlorination of" is the modification.

The *reference* enables the index user to find the information from which the index entry was derived. In the above example, 3148*i* is the reference. The reference to a deck of punched cards will be a code number representing the holes that are to be used for sorting.

An *index entry* is the assembly consisting of *heading*, *modification* (which may be lacking), and *reference*, or cross reference.

Selection of Headings for Subject Indexes

The selection of concepts or subjects that are to constitute the headings in a subject index is an art and science in itself. It requires much experience and a wide general knowledge of the subject matter. It has been found best to select index headings during the process of indexing rather than to conceive of them as an abstract mental exercise. In this way they will correspond more closely to actual usage, which, as pointed out above, helps to produce headings at places where index users are likely to look first.

Once a heading is started, it should be used consistently; this requirement puts a burden on the indexer, who either must remember the headings or must check previous indexing to find how the same material has been indexed before.

The selection of headings that will serve as efficient guides to the material indexed is the first consideration, both in time and importance. For a large index this is an exacting task which requires much experience for effective accomplishment. For every subject or phase of a subject it is best to evolve rules which must be followed fairly rigidly in order to secure a completed index free from inconsistency. The person who wishes to build an index might do well to take the headings from a good published index and use these or modify them to suit his needs. He can also use the cross references found as a start for his system. The rules for best selection of headings for any given general subject are so numerous and elaborate that it would be impossible even to list them here. A few general rules, together with the use of a good published index, will help the new indexer get started. The first rule has already been mentioned above: Put entries where index users are likely to look, even if the place selected seems illogical. Except for very general entries, do not select headings which are too broad. Thus, in an index on dyes, the heading *Dyes* would be reserved for very general entries on the subject; otherwise all of the entries in the index might go under this one heading.

In the selection of headings it is important to pay less attention to the

words used than to the subjects considered. The gist of most scientific articles can be expressed in surprisingly few sentences. If these sentences contain commonly used words and phrases there is a very good chance that these same words and phrases should be used in the index as headings and modifications.

Selection of Modifications for Subject Indexes

Once the headings have been selected from the sentences containing the gist of the material to be indexed, the selection of the modifications will, in many cases, become almost automatic. If a scientific article is about "electroplating on aluminum with copper," the modification under the heading *Copper* would naturally be "electroplating with, on Al." Modifications should be as specific as possible without loss of information. It is, of course, better that they be too broad than too narrow.

In general, it has been found most satisfactory to avoid logical classification of the modifications under most headings. Modifications are simply arranged alphabetically, with the first preposition, if any, ignored. The most important word is brought to the front of the modification where possible.

The length of the modification can be highly variable. Good modifications may consist of the whole title of the article or of a phrase taken from it or from the body of the article or abstract, or the needed information may be expressed in the indexer's words (often desirable). Good modifications may consist of a single word, or none at all may be used. There is advantage, naturally, in the shorter modifications since these save reading time. In an index, one is often astonished by how much can be said in five or six words. Short modifications may indicate broad treatment of the heading; the use of no modification at all may indicate the broadest possible treatment, or that the subject expressed by the heading is narrow or specific. The philosophy behind the omission of a modification for an unusual or specific subject is that a person searching for its heading is no doubt going to look up all entries regardless of modification. This situation occurs usually for headings which are the names of rare chemical compounds.

Assembly of the Index

To the amateur indexer, this step in index building seems to present most mental hazards; to a professional organization set up for indexing, this is one of the fastest and least expensive steps in the whole process.

If the index is to be very small and without modifications, it can be written in a notebook with plenty of space left for additional headings. A looseleaf notebook will give increased flexibility. For larger indexes some sort of cards can be used to greatest advantage, with one entry per card. If the final index is to be typed or printed from the cards, these can be of

20-pound bond paper rather than of the heavy library-type card material. This is a great drawer-space saver since these cards will run about 250 per inch. Even if the card index is to be used directly, cards made of paper may be satisfactory, for experience has shown that bond paper is surprisingly durable. A 3- by 5-inch card is a convenient size. If the cards are to be shipped in bundles a centered hole near the bottom edge will facilitate tying small bundles with string to prevent mishaps leading to disorder.

It is satisfactory either to write on the cards in longhand or to type them. If they are of bond paper, rather than card stock, typing is simplified. The heading may be typed in the upper left corner, the modification in the center, and the code number or reference in the lower left corner. Such cards can then be conveniently arranged by alphabetizing the headings, arranging the modifications (if there are any) in the desired order, and putting the references of cards with like headings and modifications in numerical order.

After the index cards have been assembled in the proper order, it will be desirable to *edit* them. The principal purpose is to harmonize both headings and modifications by getting like ones together. It may be decided that some headings are too broad; that is, if kept complete, the heading would accumulate so many entries as to be unwieldy. An example of such a heading would be *Colloid chemistry* in a subject index to the whole field of chemistry. If every entry dealing with the subject "colloid chemistry" were put under this heading, the number of entries would be so large as to make the heading unusable. It may be found desirable to restrict such broad headings to entries of a very general type, as for a review or a historical treatise.

Mechanics of Indexing

The mechanics of indexing is comparatively simple. Consider how index cards are made for a standard index to the literature. The indexing can be carried out in several steps if this seems most efficient. The first step is that of marking. In this, the word or phrase serving as the index heading is underlined on the pages to be indexed, or, if not present, is written in. The second step is that of card making. An indexer takes the marked copy, types or writes the underlined word or phrase at the top of an index card, invents a modification and puts it on the middle of the card, and enters the reference in the lower left corner. It has been found more efficient to combine the marking and card-making steps, provided that sufficiently skilled help is available. For certain purposes it may be sufficient to have a technically trained person mark (this task calling for higher technical skill) and have a person with less technical skill write the cards, including modifications.

After the cards are made, it is well to have another technically trained

person check them against the marked copy to see if more cards should be made, if the modifications can be improved, if some cards can be eliminated, and if any errors have been made. The reference can be independently checked most efficiently by less technically trained help. In some cases it may be sufficient to check the references merely for consecutiveness.

After the checking operations, the cards are alphabetized or arranged into a classification, depending on the type of index.

The assembled cards are then edited, including application of the cross-reference and inverted cross-reference systems.

The cards are then ready for use, usually after being printed. Indexes to *Chemical Abstracts* are printed directly from cards with careful checking and rechecking for the sake of accuracy.

Index Searching

The qualifications of a good indexer are likewise good qualifications for the subject-index user. The use of subject indexes is more of an art than is generally realized. Effective searches require special knowledge, training, experience, and the exercise of judgment on the part of the searcher. He must draw heavily on his general fund of knowledge, must know what to expect of subject indexes, and must go about the task with the same thoughtful and alert attitude that is appropriate when one is seeking information in the laboratory. The difference between success and failure in finding a single bit of information or in making a reasonably complete general search may lie not so much in the indexes as in the index user.

For good results the index searcher must meet the index maker part way. The using as well as the building of indexes is an art. Conscious effort to become a good index user will repay any scientist. It is first necessary to become familiar with the characteristics and peculiarities of the various kinds of indexes which one may have occasion to use. While the indexes to *Chemical Abstracts* are not dependent for ready and effective use on the Introduction which is provided (an index should stand on its own feet), the serious searcher makes a mistake if he does not study his indexes. The Introduction, just mentioned, is intended to familiarize index users with the indexes to *Chemical Abstracts* and to help them gain the maximum advantage from their use.

It is important to avoid being too easily satisfied in the use of an index, and persistent effort is likely to be rewarding.

The place to begin looking in an index is under the heading coming first to mind. It is not wise to try to "outwit" the indexer. The good indexer's first concern is putting the information where the user will look first, or at least in providing a cross reference when some special consideration, as standard or systematic nomenclature, dictates something different. If the first heading searched fails to disclose the information, then an array of

search places can be made. The array is probably best started with synonyms, as vitamin B₁, thiamin(e), aneurin(e), etc. When the index is searched for synonyms, "See" cross references may be found. These will lead to the headings of main interest. The remaining synonyms should be checked to eliminate the possibility of losses from scattering. This is especially true for "synonyms" which are trade names. After the main headings have been located, the array can be expanded to include more general headings, more specific headings, and otherwise related headings.

Confidence in the completeness of the array can often be increased by noting the growing frequency with which the same paper is picked up through different parts of the array. The array should expand with the literature search and also with the laboratory work. Index searching is a developing, unfolding sort of process; the more one knows, the more varied is the information that becomes acceptable and useful. It is highly improbable that all of the useful information will spring from the index at first glance.

It is usually best to make one's own searches. The personal element in index searching and the possibility for growth in a search through applying one's self to the task raises a question as to the loss in this respect in the use of punched cards. Though for many purposes punched cards may prove to have more than compensating advantages, there is no denying the fact that that part of the literature search which is contributed by the searcher himself as contrasted to the indexer in the various ways discussed above may prove a serious loss to the individual who might depend wholly on mechanized systems.

Once a desired heading has been located, the best method of searching among the modifications becomes the next concern. Users of mechanized systems may not be faced with the problems involved in locating significant modifications. For indexes with modifications, a glance will often reveal the desired information or its absence. Larger headings require scrutiny. Since it may frequently be difficult to predict the word starting a modification, it may be necessary to read every modification. This is not such a source of lost time as might at first be supposed since substitute, related, and accidental information may be discovered. The modifications can be arranged into order of probable importance to the index user, or a number of modifications can be selected and arranged likewise. The list can be looked up, starting with the most important, and the process continued until too little relevant information is discovered to justify continuance. After the main entries have been looked up, read, and digested, it is sometimes advisable to re-read the modifications under the more important headings since the increased background of knowledge on the subject will make more information understandable and acceptable to the searcher.

Now that a start into the literature has been made, the references found

printed in most scientific papers will probably provide much further information. Author indexes are often helpful at this point in locating more information since authors tend to specialize in their research work. In some respects, searching the literature may resemble a nuclear chain reaction; it may be difficult to start, but, once started, it may be difficult to control since the reference-multiplication factor is usually much greater than one.

It is desirable to read the entries under a heading slowly because it takes time to include the supplementary information that the indexer had to leave out of the index. One might think of an index user as a paleontologist. The entries he discovers are the "bones" of yesterday's information. He must judge, from their size, shape, quality, and location, to what kind of subject animal they belonged. The paleontologist reconstructs the whole animal from the fragment of bone he holds in his hand; the index user reconstructs the source information, with similar imagination, from the entry he holds in his mind.

Conclusion

The obstacles in index building seem formidable to the beginner. If the principles given in this chapter are followed, few big mistakes should result, at least after experience has been gained. The main things to remember are: Summarize the information to be indexed, or some unit of it, into a few sentences expressed in commonly used terms carefully chosen. Select one or more of the important words in these sentences as headings for the index. Make the modifications as specific as possible without loss of information. Once a heading has been started, maintain it consistently unless it turns out to be inappropriate. Strive for the maximum specificity in headings (i.e., do not index butene under *Olefins*, unless the index is to be a classified one). Above all, keep in mind that an index is a guide to information.

An index can be used more effectively by constructing an array of search places consisting of synonyms, more general, more specific, and otherwise-related words. This array should expand with both the library and the laboratory work.

Mechanization of parts of the chemical literature is an accomplished fact; mechanization of all of the literature in such a way that it will be available to most chemists is a possibility. This possibility is being investigated.

Supplementary Reading

- Baker, Catherine, "Making an Index," *Public Libraries*, 10, 132-35 (1905).
Bernier, C. L., and Crane, E. J., "Indexing Abstracts," *Ind. Eng. Chem.*, 40, 725-30 (1948).
Blair, E. H., "Indexing and Indexers," *Am. Library Assoc. Bull.*, 5, 234-38 (July, 1911).

- Brooks, A. A., "What I expect of an Index," *Library J.*, 35, 51-55 (Feb., 1910).
- Brown, G. E., "Indexing, a Handbook of Instruction," London, The Coptic Series, 1921.
- Clarke, A. L., "Manual of Practical Indexing," London, Library Supply Co., 1904; "Reform in Indexing Methods," *Library World*, 9, 317-20 (Mar., 1907).
- Crane, E. J., Patterson, A. M., and Marr, Eleanor B., "A Guide to the Literature of Chemistry," 227-85, 1957.
- Fletcher, W. I., "Some Points in Indexing," *Library J.*, 4, 243-47 (July-Aug., 1879).
- Fulcher, G. S., "Indexing of Scientific Articles," 1922.
- Hetherington, E., "The Indexing of Periodicals," *Index to the Periodical Literature of the World*, 2, 3-7 (1892).
- Nichols, J. B., "Indexing; Manual for Librarians, Authors, and Publishers," *Publishers' Weekly*, o.p., 14 (1892).
- Perkins, F. B., "Book Indexes, U. S. . . . Education, Bureau of. Public Libraries in the U. S.—Their History, Condition, and Management," *Special Rept.*, Pt. 1, 727-32 (1876).
- Pollard, A. W., "Making of Early Indexes. Interesting Historical Sketch," *Living Age*, 257, 29-36 (Apr. 4, 1908).
- Robinson, O. H., "On Indexing Periodicals and Miscellaneous Literature, U. S. . . . Education, Bureau of. Public Libraries in the U. S., Their History, Condition, and Management," *Special Rept.*, Pt. 1, 663-72 (1876).
- Shepardson, G. D., "Suggestions for an Index of Engineering Literature: With Discussion," *Am. Inst. Elec. Engrs. Trans.*, 11, 777-83, 835-59 (Nov., 1894).
- Thomson, W. S., "Practical Guide to Indexing and Précis Writing," 289, D. Aberdeen, 1891.
- Warren, I., "Filing and Indexing with Business Procedures," 1924.
- Wheatley, H. B., "How to Make an Index," 1902. "What is an Index?" London, Index Society. o.p. 1879.
- Wheeler, M. T., *Indexing Bull.*, No. 779 Albany, University of State of New York (May 1, 1923).
- Wilson, H. W., "Periodical Indexes," *Library J.*, 38, 591 (Oct., 1913).
- "Encyclopaedia Britannica," 12, 148 (1946).

Nomenclature Pamphlets

(Available for distribution from the office of the Chemical Abstracts Service, The Ohio State University, Columbus 10, Ohio.)

1. *Definitive Report of the Commission on the Reform of the Nomenclature of Organic Chemistry*. Translation with comment and index by Austin M. Patterson. 20 cents.
2. *Rules for Naming Inorganic Compounds*. Report of the Committee of the International Union of Chemistry for the Reform of Inorganic Chemical Nomenclature, 1940. The committee has provided an American version of the rules. (Out of print. Revised rules will be available in 1959).
3. *The Pronunciation of Chemical Words*. A committee report. 5 cents.
4. *Nomenclature of the Hydrogen Isotopes and Their Compounds*. A committee report. No charge.
5. *Directions for Abstractors and Section Editors of Chemical Abstracts*. Much concentrated information on nomenclature, symbols, forms, and abbreviations is assembled in this 46-page booklet in form convenient for use. 25 cents.
6. *The Standardization of Chemical Nomenclature*. This reprint of an article by the

committee chairman contains a list of references to sources of information on chemical nomenclature. No charge.

7. *The Naming and Indexing of Chemical Compounds by Chemical Abstracts*. Introduction to the 1945 Subject Index. A comprehensive, 109-page discussion of chemical nomenclature as applied to inorganic as well as organic compounds for systematic indexing, with a classified bibliography, an index, and the following appendixes (lists): (I) Miscellaneous chemical prefixes, (II) Inorganic groups and radicals, (III) Anions, (IV) Organic groups and radicals, (V) Organic suffixes, and (VI) 1945 ring index. Included as an insert of the 6-page 1947 Subject Index Introduction to show changes and additions. The 1952 Introduction also is included. \$1.00.

8. *The Nomenclature of the Carotenoid Pigments*. Report of the Committee on Biochemical Nomenclature of the National Research Council, accepted by the Committee on Nomenclature, Spelling and Pronunciation of the American Chemical Society. No charge.

9. *The Naming of Cis and Trans Isomers of Hydrocarbons Containing Olefin Double Bonds*. A committee report. No charge.

10. *The Designation of "Extra" Hydrogen in Naming Cyclic Compounds*. A committee report. No charge.

11. *Avogram*. A committee report. No charge.

12. *The Naming of Geometric Isomers of Polyalkyl Monocycloalkanes*. A committee report. No charge.

13. *Commission de Nomenclature de Chimie Inorganique*. This relates to the names of new elements or others concerning which there has been controversy as to names. In English. 10 cents.

14. *Commission de Nomenclature de Chimie Biologique*. Report on the 1955 meeting at Zurich (in English). This includes rules (part of them tentative) for the nomenclature of vitamins and tentative rules for steroid nomenclature. 75 cents.

15. *Commission de Nomenclature de Chimie Organique*. This includes rules on the nomenclature of organosilicon compounds (now covered by Item 22 below), changes and additions to the definitive report, extended examples of radical names, and an extensive list of radical names. All in English. 50 cents. Amsterdam, 1949.

16. *Commission de Nomenclature de Chimie Organique*. Report on the 1955 meeting at Zurich (in English). This includes tentative rules for the nomenclature of acyclic and cyclic hydrocarbons and heterocyclic compounds. \$1.00.

17. *Arene and Arylene*. A committee report. No charge.

18. *Halogenated Derivatives of Hydrocarbons*. A committee report. 10 cents.

19. *Use of "Per" in Naming Halogenated Organic Compounds*. A committee report. 10 cents.

20. *Use of "H" to Designate Position of Hydrogens in Almost Completely Fluorinated Organic Compounds*. A committee report. 10 cents.

21. *Organic Compounds Containing Phosphorus*. A committee report. 25 cents.

22. *Organosilicon Compounds*. A committee report. 10 cents.

23. *Nomenclature of Natural Amino Acids and Related Substances*. A committee report. 25 cents.

24. *Carbohydrate Nomenclature*. A committee report. 25 cents.

25. *A New General System for the Naming of Stereoisomers*. Rules proposed by G. E. McCasland and considered promising by the Advisory Committee on Configurational Nomenclature, but not official now. The advisory committee's first report is also included. 50 cents.

26. *Introduction to the 1952 Subject Index*. This includes an extensive list of organic groups and radicals (see No. 7 above). 25 cents.
27. *Nomenclature for Terpene Hydrocarbons*. A committee report. 25 cents.
28. *"Petrochemistry" and Its Variants*. A committee report. 5 cents.
29. *Commission des Symboles et de Terminologie Physicochimique*. Report of the 1955 meeting at Zurich (in English). This includes many standardized symbols for units or quantities. 75 cents.

Chapter 25

MAKING CLASSIFICATION SYSTEMS FOR PUNCHED-CARD CODING

NORMAN T. BALL

*Inter-Departmental Committee on Scientific Research and Development
Washington, D. C.*

Introduction

Technical information in quantity is often easier to handle if it is sorted into organized and related subject groups. This fundamental truth has been recently emphasized with the adaptation of punched cards and other mechanical and electronic devices to the sorting and selecting of desired groups or units of information from an organized collection. The term "classification" is used in this discussion to mean the development of systems for grouping together like items of information in an orderly and useful arrangement. By the judicious application of a numerical or alphabetical code to such a system of related group titles it is possible to extract any given piece of information or the paper or book containing it from the mass. By various devices, from the hand-sorting needle to the more complex mechanical sorting or collating machine, we can then enter the system by means of the code and move rapidly through the pattern to isolate the information which satisfies our requirements.

This *code*, however, which is so useful in applying machine methods, must be recognized as having its existence apart from the classification system to which it is applied. Coding may be likened to the attachment of a handle to a bulky or awkward physical object. *Coding* may be described as the assignment of a brief and simple pattern of symbols, such as letters or numerals, to represent longer and more cumbersome words and phrases which define the scope of a particular class. Such brief symbol patterns should have an expressed or understood descriptive definition in order to assure that the class content is uniform and understood by all who use the system.

Frequently, a code is applied to concepts in a purely random arrangement. This is common in many cryptographic schemes used to convey messages in secret. Punched cards may be employed with any code, however random, by the use of an interpreting code book which defines the meaning of each code group. The presence of a code notation does not in itself necessarily mean that the concepts coded are systematically arranged or classified.

A *classification system* worthy of the name is organized in such a way that the material is not only broken into relatively homogeneous groups, but the groups are arranged so that relationships between various groups are apparent upon scanning the schedule of class titles. The most obvious of these arrangements, of course, is that of the whole of a concept to its parts. It is usually assumed that the classifier has reviewed the entire scope of the field which is being classified, and has divided it, first into its major categories and then into further subdivisions adding up to the entirety of each of such major categories. These subdivisions are groups of individual items having common characteristics which permit them to be considered together. There are obviously as many ways of grouping like items as there are aspects of similarity. The relative usefulness of any classification system will be found to be much enhanced if a consistent point of view is maintained in selecting the characteristics used to determine likeness. For example, chemical compounds might be classified according to structure, properties, or uses. The characteristics that are used for making the groups should be chosen for their pertinence to the type of work for which the classification system is being made. It would be pointless, for example, to collect items of scientific detail according to the address of their writers unless, of course, the purpose were to make it possible to discover sources of such information on a purely geographical basis.

The divisions of a classification system will be assumed to add up to the total subject matter being classified. Here again, the use intended for the system must be taken into consideration. The effectiveness of a scheme will be weakened if it tries to cover a broader field than is involved in the problem attacked. Each division, in turn, is made up of a number of subdivisions adding up to the entirety of that division, and so on down to more minute subdivisions as may be necessary or desirable. It has become customary in applying code notation to such divisions to indicate the degree of indentation by the number of digits of the code applied to a given degree of indentation, as for example in the Dewey Decimal System:

5 (00)	Pure science
51 (0)	Mathematics
52 (0)	Astronomy
53 (0)	Physics
54 (0)	Chemistry
546	Inorganic
547	Organic

Such a notation has a special value with punched cards in that it makes possible a selection of large or small groups of punched cards according to need. If all cards relating to "Pure Science" in the above system are required, we merely select, either by hand or by appropriate machine, all

cards having the first digit —5— punched or slotted on the card. If that brings out too great a number, the further requirement that we want only that portion of pure science relating to "Chemistry" is made by selecting the second digit —4—. To go still further and specify "Organic Chemistry" we add the third digit —7—.

The selection of —54— would have included the 547 group of "Organic Chemistry" and, in addition, all other items under "Pure Science" which were classified as "Chemistry." The effect of the third digit —7— of the 547 is to eliminate from consideration all fields of chemistry in which there is no immediate interest. From this standpoint, it may be seen that effective classification methods make it possible to discard with rapidity and precision great masses of material as being outside our momentary field of study.

Most accounting machines have, of course, been devised for dealing with numbers, and therefore operate well with codes arranged in a decimal relationship. Often simple changes will adapt them to codes using the letters of the alphabet. This permits the subdivision of categories at each level into as many as 26 groups. This is mentioned because a rigidly enforced decimal breakdown of subclasses is sometimes inconsistent with a logical grouping. A scheme which has become very popular because of its flexibility is known as the "punctate system." This separates successive breakdowns by a point, as in the decimal system, but permits more than one digit, if required, between points. For example, (2.8.12) would indicate the 12th subdivision of the 8th subclass of the 2nd class.

A great advantage of the punched card over previous methods of manipulation and selection of material lies in its ability to carry a number of code designations arranged on different portions of the card. These may indicate different subdivisions of a single system, thus providing for accurate location of several contributing parts of an item. The important point, however, is that they may indicate appropriate subdivisions in entirely different systems of classification, divided according to entirely different criteria to accomplish different purposes. This is most effectively accomplished with a number of quickly devised relatively simple systems. Each one must have a single clear purpose and should endeavor to meet the other requirements of effective classification set forth later in this chapter. One system may be subdivided according to the structure of a given compound, while another may indicate the function for which the material is used. This gives us a method, similar to that of the simultaneous equation in algebra, whereby we can solve a complex problem in information research in which we recognize a number of independent variables.

A very simple example may make the point more clear. Suppose that we are to classify automobiles. We may have Chevrolets, Fords, Plymouths, Buicks, etc. An individual car may be new or used, have eight

cylinders or six cylinders, and an almost endless variety of colors. By the old-fashioned methods, using paper and pencil alone, we are forced into some such laborious pattern as follows:

CHEVROLETS

New

8-cylinders

Blue

Gray

Yellow

6-cylinders

Blue

Gray

Yellow

Used

8-cylinders

Blue

Gray

Yellow

6-cylinders

Blue

Gray

Yellow

For Fords, Plymouths, Buicks and other makes of cars, we require a similarly extensive system.

Here, of course, we are faced with the necessity of choosing the criteria to be used for major classes and those to be subordinated. As the elements within each group become more numerous, the pattern becomes more involved. Usually, there are many more than two possible terms in each variable. The complexity of such problems has been largely responsible for the despair with which many people discuss classification as a technique. With punched cards, and similarly with the more complex mechanical and electronic machines, a number of independent variables may be sorted simultaneously. On punched cards, our automobiles may be classified as follows and coded in four independent fields:

1. (make) Chevrolet
Ford
Plymouth
Buick
2. (age) New
Used
3. (motor) 8-cylinders
6-cylinders
4. (color) Blue
Gray
Yellow

Our sorting manipulator is then set to select those cards on which the desired characteristics are satisfied in each field. This technique is applicable to any system using punched cards for sorting or for selecting fields of search. It applies equally well to the more complex mechanical and electronic devices, such as described in Chapter 3.

It should be recognized that there is no uniform proper order for the arrangement of classes and subclasses. In discussing this point with two of the editors of this book some interesting actual experiences were described. They illustrate variations in arrangements which are caused by differences in points of view. One man had developed a system for classifying bibliographical material pertaining to "writing ink," one of the principal fields of interest of his company. He found it useful to collect a subordinate group under the title of "dyes." The other editor faced a different problem. One of the principal products of his company was "dyes." It was appropriate, to him, therefore, that this should be a major category. In such a system the term "writing ink" was a relatively minor subdivision providing for a specific application of dyes. A third arrangement would be appropriate in a listing of materials to be purchased by a small manufacturing concern. It is possible that the relative size and relationship of requirements would justify having "dyes" and "writing ink" as equal groups in planning or accounting for the expenses of the plant.

Collateral Values of Classification

A subject-matter classification system which is tailor-made to fit the type and quantity of information to be classified is especially useful to determine how well a given field is covered by available information. The main titles, and their subdivisions, to whatever extent required, serve as an outline of the information classified. A number of items having a common subject require only a single title in the classification outline, and can be considered as one item. Such an outline can readily be made to show gaps and duplication of effort, and often suggests new paths for developing an association of related ideas and relationships between fields commonly accepted as separate. The breaking down of large fields into their component subdivisions which show up clearly in the classification outline is particularly effective in providing a ready selection of trails into a large wilderness of materials. The discovery of adjacent concepts arranged in proximity and the listing of various parts of a concept in their relation to the whole can stimulate the inquiring mind to broaden the scope of an investigation and accomplish its purposes on a much larger scale than anticipated in the original approach.

Classification has no substitute as a method of organizing material for determining exhaustively (e.g., *all* material that is related to a given

subject). The Patent Office, for example, must have a system whereby similar and related concepts are classified to show relationships. The right to a patent depends on the *fact* of novelty. The breadth of the claims depends on the extent of that novelty. A search, therefore, must first of all determine with some precision whether or not the invention has been made previously. Then the exact point of conflict with related previously disclosed ideas must also be ascertained in order that the breadth of protection to be granted the inventor by the Patent Office may be accurately determined.

A cogent statement of the value of classification is given by Stanley Jevons:

"Science . . . is the detection of identity, and classification is the placing together, either in thought or in actual proximity of space, of those objects between which identity has been detected. . . . Whenever we form a class we reduce multiplicity to unity. . . . Of every class, so far as it is correctly formed, the principle of substitution is true, and whatever we know of one object in a class we know of the other objects, so far as identity has been detected between them . . . it is the exertion of the classifying and generalizing powers which enables the intellect of man to cope in some degree with the infinite number of natural phenomena"¹.

The making of any classification system is a time-consuming operation. It requires a high degree of mental application in cutting across confusing terminology and determining the most useful generalizations. A schedule of subject matter arranged without a serious effort to juxtapose similar groups and to relate larger concepts to their constituent parts as genus and species will rarely give the results that are expected of classification systems. Experience has shown that poorly prepared classification systems, especially when intended as the basis for a punched-card system, can be far less effective than simple alphabetical indexes. Professor Toops of Ohio State University, for example, has found that the very speed and accuracy of machine operations tend sharply to point up faulty planning².

The Alphabetical Index

The alphabetical index is often a completely adequate method of finding items of information in a larger mass. This is especially true where nomenclature is well accepted and standardized, as it is in many fields of chemistry. It is also adequate if the contemplated search involves only the findings of representative examples of a given subject matter without the need for certainty that the search is exhaustive, or the necessity of finding

¹ Jevons, W. S., "The Principles of Science," 2nd Ed., 673-4, New York, The Macmillan Co., 1900.

² Baehne, G. W., "Practical Applications of the Punched-Card Method in Colleges and Universities," 177, New York, Columbia University Press, 1935.

related material. It becomes less effective as the mass of material becomes larger and begins to involve a wide range of terminology.

An alphabetical index is a very precise type of classification. The criterion used in setting it up is the relative arrangement of letters in the alphabet. This is a clearly established pattern. If the words used in selecting titles or headings in an index are accurate and well standardized, they provide all the information necessary to find an item having a title or heading as listed.

The dangers in complete reliance on indexing for exhaustive search purposes lie in the practice of widely scattering the references to similar subjects having names spaced widely apart alphabetically. "Applications" would appear at the beginning of such a list, separated from "uses" by many unrelated terms. It is not humanly possible to remember all words relating to a given area of scientific effort, especially when dealing with indexes with thousands of entries, as would be involved today in any large-scale cataloging of science or technology or even of any one major science or branch of technology. A system of complete cross-referencing among all related entries in a large index creates an unwieldy bulk and a sense of frustration in trying to follow a train of association.

However, it must be clearly pointed out that almost all classification systems of any size or complexity recognize the need for alphabetical indexing to the class titles as an integral part of their method. For one thing, this practice aids in maintaining a consistent point of view in making later decisions between apparently conflicting possible locations of material. The Dewey Decimal System has its *Relative Index*. The Library of Congress uses, with its very complex system of classification, an elaborate card index merging a number of possible avenues of entry to the system. The same card catalog contains entries for author and title as well as for subject matter. The Patent Office Classification Manual has an elaborate and extensive index, which is kept in a growing state by the constant interpolation of punched cards in a cumulating stack as new titles are suggested. They serve as additional clues in reaching a desired part of the classification outline for starting a search. Contributions to this index from users inside and outside of the Patent Office are actively encouraged. All difficult trails located as a result of consultation with the Classification Division of the Patent Office are added to the index in every way possible in the form of additional punched cards. The advantages of punched cards in preparing indexes are worth a mention here. Dean Fletcher of the University of Illinois discovered that punched cards could save as much as 80 per cent of usual time spent in sorting and arranging index cards³.

³ *Ibid.*, 407.

Quick Discard of Unwanted Material

There is one very striking advantage to the practice in a classification system of dividing material into a few major groups, each in turn subdivided into smaller units. We discard, almost instantaneously, great portions of the mass of material being attacked. Only one decision is required at each step toward the detail sought. Ten-thousand headings arranged in decimal fashion might under ideal conditions (equal distribution of material under each heading) require only four decisions to determine precisely the point of search. The first decision eliminates nine-tenths; the second decision discards nine-tenths of the original tenth selected, leaving only one-hundredth of the original mass for further consideration, and so on through the pattern. The fourth choice leaves only $0.1 \times 0.1 \times 0.1 \times 0.1$ or one-ten-thousandth of the original collection for further attention. This is a spectacular difference from the slower progress through a long alphabetical listing. Nothing is really eliminated from attention at any step from word to word except the single item being considered for selection or discard. This accomplishment, whereby we eliminate by single decisions large percentages of nonpertinent material from consideration, is one of the fundamental objectives of subject classification. In no other way, for example, would it be possible to attack masses having millions of items, as is done in making a patent search.

To approach this ideal, however, it is necessary to accept and follow a few principles, discovered by experience to be important. Sometimes the attempt to satisfy them all will not be possible. However, if they are considered one by one in any endeavor to make a classification scheme and discarded only if found to have little value in the problem at hand, they will serve to avoid troublesome pitfalls. In discussing "Problems of Classifying Chemical Patents," Mr. M. C. Rosa of the Patent Office expressed the real test clearly:

"The main desideratum in the mind of the classifier is to evolve a classification which will do the most good to the greatest number of searchers. He soon learns that a perfect classification or one which will prove satisfactory to every search is impossible of attainment and strives, instead, to produce a classification that adequately provides for the subject matter it embraces and is flexible enough to accommodate innovations . . ." ⁴.

Requirements of a Classification System

Purpose. The first requirement for the establishment of a workable system of classification is a clear concept of the purpose it is to serve. With

⁴ Rosa, Manuel C., "Problems of Classifying Chemical Patents," *J. of Patent Office Society*, **24**, 241 (Apr., 1947).

this must be coupled a willingness to omit and exclude subdivisions which do not contribute to that purpose. It is much more effective to have several simple and straightforward systems for achieving clearly defined objectives than to labor over an attempt to organize great segments of human knowledge for a number of conflicting purposes. Although it may seem an unnecessary burden at the time, the pains of preparing a written expression of the scope and purpose of the system will always be repaid by clear conceptions and easier working decisions. However obvious the purpose may seem, its limitations become vague with the passage of time and may be confused in the consideration of complex details. There are frequently many possible ways of classifying a given number of items. In broad subjects, such as chemistry, for example, the number of alternatives may approach infinity.

According to the introductory pages of the British Standards Institute Translation of the Universal Decimal Classification:

"Many classifications, of knowledge in general and of the sciences in particular, have been evolved rather as an intellectual exercise than with any specific aim in view. Numerous attempts have been made to devise so-called logical classifications, designed to satisfy the ideal criterion that the order of sequence of the classes shall correspond with the logical order of sequence of the concepts represented. It is assumed that a logical or 'natural' relative order of the various branches of human knowledge and activities exists and can be determined. That this assumption is false and the pursuit of the ideal illusory, is demonstrated by the rapidity with which such attempts have lapsed into oblivion. On the other hand, classifications which have initially been designed to meet a specific practical purpose are nearly always found to have a reasonably logical sequence, and to be readily adaptable to other purposes, provided these latter are not too remote in characteristic from the original function for which the classification was designed" ⁵.

This may be based to some extent on the more abstract words of John Dewey:

"Organization is no more merely nominal or mental in any art, including the art of inquiry, than it is in a department store or railway system. The necessity of execution supplies objective criteria. Things have to be sorted out and arranged so that their grouping will promote successful action for ends. Convenience, economy, and efficiency are the bases of classification. . . .

" . . . They must be arranged so as not to overlap, for otherwise when they are applied to new events they interfere and produce confusion. There must not only be streets but the streets must be laid out with reference to facilitating passage from any one to any other. Classification transforms a wilderness of byways in experience into a well-ordered system of roads, promoting transportation and communication in inquiry" ⁶.

⁵ *Universal Decimal Classification*, 1, 3, London, British Standards Institution (Jan., 1943).

⁶ Dewey, John, "Reconstruction in Philosophy," 154, New York, Henry Holt and Co., 1920.

Main Divisions. The first step in evolving an actual schedule of classification groups is the selection of the point of view on which the main divisions will be based. For maximum effectiveness there should be only one clear basis of major division. Here again, easier working decisions come from a clear statement of the criteria to be used in distinguishing between these main divisions. In practice, some compromise may be acceptable. For example, library systems have subdivisions based on *form* of material in the same schedules with others based on subject content. Dewey, for example, places scientific dictionaries in a minor subdivision —503—. Later subdivisions under —500— are made on the basis of the particular science or branch of science forming the subject matter. Physics is provided for in the subdivisions arranged under —530—, and chemistry under —540—. Such compromise will do little harm if it is recognized as a definite exception to a rule. It is very harmful if made at random and without clear purpose.

The sorting of existing physical objects is relatively easy. It is little trouble to group together objects which look alike, or those which work together to perform a common purpose. Dishes or other household articles are commonly arranged in similar groups for everyday convenience. Many of us have elaborate systems for sorting nails and screws by size and other standard characteristics to permit instantaneous selection of the one to fit a particular job.

As we pass from the physical object to its mental concept, the problem becomes less simple. There is still a relatively happy solution if we are able to deal entirely with existing and visible material to be sorted into categories in which similarity is apparent.

Often, however, we are dealing with partial samples of material which is to appear or be collected in quantity in the future. In such cases we must use those samples to generate a mental concept of the most inclusive expression describing that type of material as we hope to collect it in our system. We must then name our groups in terms that will not exclude a forgotten item. This last factor is often overlooked in the eagerness to use the more specific name at hand, which applies only to the limited characteristics of the particular samples. Effective *generalization* is a very useful mental operation, and is half the secret of good classification technique.

Hospitality. A scheme of classification must provide a home for all material contributing to the purpose for which it was devised. This seems an obvious statement, but it is a characteristic often neglected. Many systems are made up of a list of groups which represent only titles in mind at the moment they were listed. By careful consideration a broad title or group name may be developed to include not only the present samples but also future additions of a similar nature.

Except in dealing with a completely collected mass of material some sort of a "miscellaneous" group is almost always a necessity. This may be

the parent group itself in a system using a true decimal notation. For example, under —540— of the Dewey system may be placed such chemical material as does not satisfy the titles of the indented subclasses. Such a group is very useful in a system in which information is to be further collected, and the direction and degree of its expansion is uncertain. As it collects, patterns will appear. The expansion and development can then proceed on empirical lines, and represent reality rather than theory.

This practice of making a miscellaneous receiving station has a further value; it eliminates the need for specified places for small or uncertain categories during the growing period. This keeps the list of groups smaller and reduces the study necessary to locate the point of a search. A classification system loses effectiveness if the groups contain too few items.

Consistency. The meaning of terms and headings used in a schedule of classification should be consistent and invariable. We must be able to rely upon the invariability of word meanings because each decision we make in choosing a classification heading when searching, discards a large portion of the material in the system, and we must be sure that what is discarded contains nothing that is pertinent to our search. A searcher may become accustomed to a certain type of arrangement and assume that it is always followed. He may then decide that there is nothing on his subject, whereas it is in reality hidden in a different arrangement. One value of classification is the fact that habits of search are rapidly formed, eliminating the repeated study and analysis of individual items in long lists of alphabetical index terms.

An extension of this factor of consistency is called *Modulation*. In scanning the subdivisions of a system, it is very helpful if the rate of breakdown is fairly uniform. For example, we may select as coordinate subdivisions of chemistry, organic and inorganic chemistry. These have substantially even rank. However, if in dividing organic chemistry into broad fields (e.g., aliphatic and aromatic), we jump directly to the elements of molecular structure as the very next step under inorganic chemistry, the "modulation" is uneven. This may be a source of confusion in directing attention to the proper field of search.

Definitions. Consistency is often best achieved by establishing definitions of terms as they are decided. It may be necessary to modify these definitions as new aspects appear. It is safer to recognize the fact that the definitions must be changed than to stretch first one class title and then another to permit the inclusion of material different from the original concept of either term.

A set of definitions is always an important addition to a classification system. It is an unfortunate human weakness that details are forgotten as time passes. The details of meaning originally given to the title of a classification

subdivision are not exceptions. Time will be well repaid when spent in preparing definitions to consult in later searches or in allocation of added items.

Mutual Exclusiveness. Discussions of classification principles often state that classes must be "mutually exclusive." It is this factor which makes it possible to discard great portions of the arrangement with so few decisions. The objective, of course, is to have one group and one only which will accept a given item. The title or description of that group should, ideally, make it obvious that the item belongs there and nowhere else. The accomplishment of "mutual exclusiveness" often requires great care in the selection of subclass titles, and sometimes, because of the nature of the subject matter, cannot be attained. Here, again, the value of careful selection of generalizing terms must be emphasized. Sometimes a definition or guiding note can make clear exactly which material is found in which subdivision. It is also desirable to have those subdivisions most nearly alike arranged in proximity. This arrangement can provide much useful guidance to simplify the decisions of the searcher and speed him toward his goal.

Cross Reference. In the common event that we cannot completely achieve this ideal of the single place for a given concept, it is well to have warning signals for the future searcher. Guiding notes attached to subclass titles may be used to direct the searcher to a more accurate point of entry to the system.

These signals should be limited to use where the division of material is more extensive than a single example or so. It is simpler in the single example to place an actual duplicate bodily in a second location. In complex situations a note of explanation in some detail attached to the class or subclass title may prevent confusion.

Predicables. The real pioneer in the analysis of the problem of classifying scientific facts was Aristotle. In his "Posterior Analytics" he names the five predicables of a given term:

Genus
Species
Difference
Property
Accident

A *genus* is a group of objects or concepts which can be broken into two or more subdivisions, called "species." The *species*, according to Aristotle, must have all the properties of the *genus*, plus something more which distinguishes it from other species. This something is its *difference*. A *property* contributes identity to a given term, and may be called a "characteristic" of the term. An *accident*, on the other hand, is some chance quality that has no necessary part in the characteristics of a term. For example, the fact

that a man has arms and legs is a property of the term "man," while the fact that his name is John is an accidental quality which has no bearing on the concept of the term "man."

These predicables are still very useful in building any classification system today. If we look for the *difference* in the *properties* that distinguish *species* from each other and from their *genus* and make sure that we are not calling an *accident* a "property," we will avoid many of the careless blunders that have made failures of so many classification ventures.

Conclusion

Finally, it is suggested that the maker of any system of classification must not overemphasize the perfection or importance of his scheme. It is safer to attack a clearly limited field. Scholars and librarians since the beginning of recorded history have spent their lives in trying to develop perfect classification schemes for all human knowledge. None worthy of that description has yet appeared, as may be proved by the continuing endeavor and by the controversy among the exponents of the various currently popular systems.

In dealing with information, as with the common physical objects with which we work, the expedient of classifying should be used when it adds to the convenience of handling or locating a given item and satisfies a real requirement. If we have only a small number of objects with which to contend, we would merely create unnecessary mental work for ourselves in an attempt to identify them with a highly complex and theoretical system of classification. On the other hand, as collections accumulate, sorting becomes necessary to avoid the inspection of great numbers of individual items. The sorting process itself is not new, but the added attention to machine methods has added a new value to the sorting process represented by classification techniques. The very efficiency of the machine has necessitated the perfecting of classifying operations. These operations will, however, always require the intellectual skills of human analysts to study and to plan the system construction and to provide "feedback" to maintain and revise the system to meet the requirements of continual change in our technology.

Recommended Reading

- Ball, N. T., "The Classification of Patents," *Special Libraries*, 38, 11-16 (Jan., 1947).
Clavier, Henri, "Some Considerations Regarding the Mechanism of Thought,"
Translated by Norman T. Ball. *American Documentation*, 6, VI, No. 4, 1955.
Fairthorne, Robert A., "The Mathematics of Classification," *The Proceedings of the British Society for International Bibliography*, Oct. 14, 1947.
Good, Carter V., and Scates, Douglas E., "The Descriptive Method. Classification,"
Chapter 5 in "Methods of Research," New York: Appleton-Century-Crofts, Inc., 1954.
Gregg, J. R., "Language of Taxonomy," New York, Columbia University Press, 1955.

- Jevons, W. Stanley, "The Principles of Science," Chapter 30, 673-734, New York, Macmillan Co., 1900.
- Kyle, Barbara, "E. G. Brisch: Something New in Classification," *Special Libraries*, 47, No. 3, March 1956.
- Perry, James W., Kent, Allen, and Berry, Madeline M., "Machine Literature Searching," New York, London: Interscience Publishers, Inc., 1956.
- Piper, R. F., and Ward, P. W., "The Organization of Knowledge," Chapter 18 in "The Fields and Methods of Knowledge," New York, A. A. Knopf, 1929.
- Richardson, E. C., "Classification—Theoretical and Practical," New York; Chas. Scribner's Sons, 1912.
- "Universal Decimal Classification," London, Published by British Standards Institute (Jan., 1943).

Chapter 26

SEARCHING THE LITERATURE

BYRON A. SOULE

Chemistry Department, University of Michigan, Ann Arbor, Michigan

A literature search is comparable with a journey. Almost every phase of the one activity has its counterpart in the other. A short trip over familiar ground requires very little planning, but an extensive tour involves many details that should be carefully arranged beforehand.

In the first place a tourist must decide exactly where he wants to go and what he wishes to see. Next come the problems of transportation and accommodations, plus the need for a guide or interpreter in foreign lands. All such obstacles can be easily removed by a travel agency with its experienced staff and world-wide connections, or the traveler may assume the responsibility himself. In this case he must expect to lose time because of unfamiliarity with the best connections, to waste effort on fruitless side trips, and possibly miss his goal entirely because of some language difficulty or failure to recognize his objective in strange surroundings. Finally, if the trip has a serious purpose, a diary is a necessity.

Individual genius is revealed not so much by independence with respect to what others have done but in the ability to see new significances, new possibilities, and new values in what has already been contributed, in short, to proceed from where others have left off. The primary objective of research is to increase knowledge. Intentional additions are made only by those who know where it should be done.

Obviously, two procedures are possible: One is to work under the direction of someone else who has located a segment of the frontier and is keeping up with its advance. The other is to study the records to ascertain what has already been accomplished and then proceed with experimental work. The value of ignorance in initial exploratory tests should always be recognized since they may not be performed if too many others report failures. One may profit or be prejudiced by the experience of others. In any case the literature is a primary research tool and takes precedence over the laboratory because published records constitute the register of past achievement and therefore determine the validity of any claim to success in meeting the objective of research.

The general acceptance of this attitude has led to the formation of societies to provide for the prompt publication and distribution of research data, and of libraries to insure the gathering and preservation of published

material. Both groups have a common purpose, namely, to make the literature available, but their guides are of necessity built on a slightly different basis. It is this difference that is responsible for some but not all of the difficulties encountered by bench chemists when they seek to consult the literature. Perhaps a brief comparison will clarify the situation.

Chemical literature has been accumulating for centuries. Today the total volume is several million pages, while the present annual increase is well over 200,000 pages. Any orderly consideration of this mass demands some classification and search aids. As a result of previous and broader experience librarians decided, late in the last century, to use two major groupings, namely, books and pamphlets, with subdivisions according to subject and form. The distinction between a book and a pamphlet is vague, but in general depends on the nature of the cover, not the contents; while subject subdivision deals with the field (Chemistry, organic) and form is concerned with the sort or type of material (journal, textbook, laboratory manual). The limiting factor in this system is the binding. While the scheme itself may be complete and detailed, in practice actual classification is restricted to the smallest division that embraces the entire book if it is not to be torn apart. Unfortunately, authors and publishers give little attention to this major problem of the library world. They feel free to include in one cover as many subjects as they wish, regardless of relationship or relevancy. Equally disturbing is the fact that librarians may consider a set of books as the unit for classification. This causes no trouble with periodicals, thanks to other guides; but, if the set is a series of monographs, like Ahrens' "Sammlung" or Adams' "Organic Reactions," when there is no contents note on the catalog card, only an examination of the individual volumes will reveal the contents after their existence has been discovered.

Obviously, a whole book is too large a search unit without a master index—an aid so constructed that, regardless of the approach, it shows the location of every item of information in every volume comprising a collection. Several hundred subject entries might be required for each one, resulting in a tremendous index; however, anything short of this goal gives the searcher the responsibility for bridging any gap between the aid needed and that given. Many special libraries have already undertaken the work for their own holdings, but for the entire scientific literature the task is enormous. Little hope can be offered for its completion in the near future unless some organization like the National Research Council assumes the responsibility.

Fortunately, several national chemical societies foresaw the difficulty years ago and took steps to alleviate it for original source material (journal articles and patents, though not books or doctoral dissertations) by the

establishment of "abstract" journals. Each of these guide services endeavors to produce, in one language, a summary of every article and patent of chemical interest, without regard to the language or country of original publication. Excellent indexes are also provided to direct a reader to search areas and information about every substance newly studied, prepared, analyzed, or discussed. The unit embraced by any reference is small since very few articles occupy more than ten pages; therefore, the coverage is detailed. In addition, the indexes furnish at least four kinds of guides: author, subject, formula, and patent number. During the early stages of a study the subject index is of major use, the formula index comes second, while the author division awaits the discovery of the names of those who have dealt with the problem.

Perhaps the most important point to grasp is that the usual guide to a library, the card catalog, covers only the books in that particular collection. It gives no hint regarding the existence or location of any others. On the other hand, abstract journal indexes are guides to the entire periodical literature for the years covered. They show what has been published and give specific references. If a journal is not available locally, *Chemical Abstracts'* "List of Periodicals Abstracted" or Gregory's "Union List of Serials" will, except in rare cases, tell where a file of the magazine is to be found. Once located, access to the desired pages is easily obtained by loan or photoduplication. Thus, in any study, anyone experienced in the art of searching can quickly accumulate the journal literature pertinent to a subject and be fairly sure of his coverage.

No such assurance is available in regard to book literature. *The United States Catalog of Books in Print*, *Cumulative Book Index*, or *Publishers' Weekly* give inadequate information as to contents and location of any volume. The Interlibrary Loan Service at any library or the Union Catalog Division at the Library of Congress can be helpful in giving name of author and title; however, if only the subject (e.g., conductometric analysis) about which information is desired is available, it is unlikely that the excellent discussion, plus extensive bibliography, in Böttger's "Physikalische Methoden" will be found. Or, again, if one does not even know that Kobe's "Inorganic Process Industries" exists, how can he find therein the chapter on the Literature of Chemistry?

The previous discussion brings us directly to a consideration of the art of searching, in which an individual intimately acquainted with the books (a person often characterized as "an ambulant catalog") is most likely to succeed in finding what is desired. Furthermore, it must be clearly understood that we are concerned only with library chemistry, the finding and gathering of what is in print. The study and evaluation of the literature for the purpose of checking data, planning experiments, and preparing a research program must be left to the laboratorian.

Before starting any literature search one should know exactly what he is looking for and as much as possible about its relationships. This orientation is perhaps most readily obtained by studying the *best* textbook or encyclopedia article on the subject and then examining the *most recent* textbook or survey. Both accounts should preferably be more general than the specific idea under consideration, for they ought to reveal what has been accomplished not only in chemistry but in related fields. For example, information on fuels for rockets and jet planes is to be found in books on interplanetary travel. These works, published in the early nineteen-twenties, were primarily for astronomers who, incidentally, along with physicists, have helped in developing the spectroscope to its present state of perfection. Similarly, our knowledge of growth hormones has also been increased by biochemists, botanists, and medical men, while certain advances in atomic structure research have been made possible by mathematicians.

In other words, the background preparation should include the perusal of work done on closely associated subjects, plus an examination of other fields for results, methods, instruments, and ideas that may be applicable to the problem at hand.

Another benefit derived from wide reading at the start is that the contributions of other investigators will reveal their efforts to put the idea into words, for it must be clearly understood that anyone making a complete search of the literature is compelled first to clothe his idea in the word or phrase that others have used to express that idea.

A limited vocabulary is one of the most serious impediments to fact finding. Consider, for example, the words that may be used to convey the idea of a place in which to live. Confined to one language a few of them are:

abode	habitat	palace	hotel	boat
house	dwelling	castle	suite	berth
home	residence	mansion	lodging	quarters
bungalow	farmhouse	villa	flat	barracks
cottage	lodge	asylum	apartment	camp
cabin	rancho	maisonette	shanty	tent
shack	hut	duplex	cave	wigwam

It is these words that must be sought when interested in one idea common to them all. An interesting example of the difficulty experienced in this phase of the work was uncovered in a search for the chemical composition of "aluminum flakes." It was found under the heading "Rubber, Compounding, Aluminum in," (i.e., by a use attack). This merely emphasizes the point that a variety of viewpoints (search headings) is essential, the angle of attack often determining success or failure in a search. Things look different when approached from different directions.

To help the searcher, aids involving words are arranged in an order based upon the letters composing the words (i.e., alphabetically). The idea itself is

not susceptible to this treatment, but the searcher is well advised to arrange his embodiments in a, b, c order as a means for saving time when consulting indexes. A list is more easily checked for all the most specific terms, including synonyms, antonyms, and trade names, as well as for the ever widening circle of more and more general terms (Bunsen—Chemists—Scientists; Biography, individual—Biography, collective; Acetaldehyde—Aldehydes, Aliphatic—Aldehydes), and for that too often baffling group, synonymous words in foreign languages, since American indexes must always be supplemented by those of other countries.

Beyond all this, every index has its peculiarities. Sometimes they are revealed in a preface, such as the introduction to the *Chemical Abstracts* indexes for 1945. More often there is no statement, whereupon the searcher assumes that the index in hand has no individuality and fails to find what he needs directly for the very simple reason that he looks in all of the wrong places first. He momentarily forgets that the M', Mc, and Mac names are arranged by preference of the indexer, not in accordance with a generally accepted rule. He overlooks the fact that some of the simplest nomenclature problems have not yet been settled for indexing purposes. FeCl₃ may appear as Iron, chloride, ferric; Iron, ferric, chloride; Ferric chloride; Eisen-III-chlorid; etc. *Chemical Abstracts* uses one order of precedence in listing the substituents of an organic compound; Beilstein's "Handbuch", 4th edition, employs another, as explained in Vol. I, page 941 of the *Main Series*. The formula index in Beilstein, Vol. 29, does not agree with those of *Chemical Abstracts*. It is true that some indexes contain cross references to help anyone unfamiliar with details of arrangement. They ought not to be considered, however, as a satisfactory substitute for an explanation of arrangement or its study.

Another element in the preparation for a search is the fact that the words used as search headings should include those that will locate so-called hidden information or less obvious relations. Every adequate report of experimental work includes a brief summary of previous findings, the reasons for the investigation, a list of the chemicals used with any necessary statements regarding quality or methods of purification and testing, a description of the apparatus with indications of its limitations, details of the calibration of all measuring devices, as well as the results obtained, necessary calculations, and conclusions believed justified by the data. The searcher's awareness of these accessories is of inestimable value, particularly in saving time. For example, if a method for the purification of some substance is desired, a quick way to find it is to look up the determination of a physical property, such as the atomic weight, boiling point, or spectrogram. If the construction, operation, or calibration of some instrument is needed, a report on work employing that instrument should

be a fruitful source of information. When a device or operation is connected with a personal name (Vorce cell, Haber process, Gabriel synthesis), a biography of the man himself may furnish sufficient data or at least indicate his period of productivity, thus limiting the time range, geography, and possibly language features of the search. The method cannot be blindly applied for it is obviously useless when the value of the Avogadro number is sought.

The direction and duration of any literature search depends entirely on the information possessed by the searcher at the outset and his success in expressing the idea in words that have been used by others for descriptive and indexing purposes. It is easy enough to bridge the gap between search aids and the object of a quest when that object is at hand, but far from simple when lack of search direction, unfamiliarity with bibliographical aids, and language difficulties intervene. One suggestion applies in every case, namely, be methodical. No quantitative analysis is undertaken without careful consideration of the method, operative technique, precautions, interferences, and time requirement. Any library investigation that is worthy of the name demands equal consideration.

Where to look for chemical information depends directly on what is wanted and why, as well as the source material at hand. Theoretically, everything that has ever been published is available, but actually no complete library exists. Geographically considered, probably the best small area in the world is the city of Washington, where the Library of Congress, together with the Department of Agriculture, Bureau of Standards, and Patent Office collections are located. A close rival is New York City. The New York Public, Chemists' Club, and Engineering Societies libraries are all excellent. Chicago technologists are well served by the John Crerar Library. Other cities, such as Philadelphia, Pittsburgh, Cleveland, and Los Angeles, can also boast of fine resources. Many private libraries, particularly those owned by the larger chemical corporations, are unquestionably superior both in holdings and staff. Our foremost universities possess extensive technical collections that are generally open to all who would use them either directly, by interlibrary loan or through reproductions.

It may be safely asserted, therefore, that library work is usually not hampered by a lack of literature. The difficulty lies in finding the page or pages upon which desired data have been recorded. Where to start the search then becomes a matter of primary importance. Too many people commence and end their studies with a cursory examination of *Chemical Abstracts*, and forget that it is a guide to, not a substitute for, the journal and patent literature and that it does not reveal the contents of books except in a very general way. Other people, equally misguided, attack every

problem by compiling a bibliography of the abstract journal references. They display great activity, but too often reproduce only what has already been well done by someone else.

Unfortunately, one of the best guides to bibliographies is now out of date; however for material up to about 1933 West and Berolzheimer's "Bibliography of Bibliographies" is very useful. Earlier lists may be found as footnotes in Fehling's "Neues Handwörterbuch" and as addenda to the various chapters in Abeggs' "Handbuch der anorganischen Chemie". More recent compilations, on a wide variety of subjects, are in *Chemical Reviews* (1924-*et seq.*) which, by the way, has two cumulated indexes, Vol. 1-10 and 1-40. For bibliographies of the work of individual scientists the finest source is Poggendorff's "Biographisch-literarisches Handwörterbuch" which includes the eminent investigators of the world down to 1940 and gives a chronological list of their literary efforts. So complete are these lists that European book dealers have been inclined to boast a little when they have chanced to find a publication not mentioned in Poggendorff. As previously stated, personal names mean little at first; but, once an investigator has been connected with a subject, everything he has written is listed in this one source and the only compiling necessary is that required to bring the Poggendorff list down to date.

Another line of approach in the initial stages of a search is through the great handbooks.

Beilstein, "Handbuch der organischen Chemie"
Grignard and Baud, "Traité de chimie organique"
Gmelin, "Handbuch der anorganischen Chemie"
Mellor, "Comprehensive Treatise"
Friend, "Textbook of Inorganic Chemistry"
Pascal, "Traité de chimie minérale"
Abderhalden, "Biochemisches Handlexikon," also his "Arbeitsmethoden"
Oppenheimer, "Die Fermente"
Heffter, "Pharmakologie"
"Handbuch der Physik"
Wien and Harms, "Experimentalphysik"

Although these reference works are more or less out of date, they can be relied upon for the time covered and the abstract journals may be used to complete the journal references.

A better way is to locate a monograph on the same or a closely related subject. For American imprints this can usually be done by searching the *U. S. Catalog*, then the *Cumulative Book Index*, and finally *Publishers' Weekly*. There are, of course, comparable aids for the books of foreign countries. While these aids cover the entire output of books, of which the chemical comprise a very small fraction, the rather broad classification at least helps in sorting out those which are scientific. In addition, a file of the

Technical Book Review Index is a very good information source. As indicated by its title, reviews of the books listed are quoted in full or in part, thus giving a more satisfactory notion of the contents of each book cited. Book reviews in and lists of "books received" by various journals (e.g., *Journal of the American Chemical Society*) can be very helpful although such hunts may become tedious. It must always be remembered that periodicals seldom secure complete coverage as they use only what is sent to them and in general do not directly solicit all material for review.

It is true that the abstract journals announce the publication of books, but too frequently offer no clue as to their contents, except inferentially by placement in a particular section. For some years the *Chemisches Zentralblatt* called attention to the book notices in each issue by printing the page reference on the cover, but no hint was given regarding the section. As the subject indexes were classified, they were a little more helpful.

Since its inception the *Transactions of the Faraday Society* has averaged at least one symposium issue a year. These symposia have covered a wide variety of topics, chiefly physicochemical in nature. Other journals have also devoted space to symposia papers, but not with the same regularity. So any search for physicochemical material should include a direct examination of the *Transactions*, without relying upon the abstract journals to reveal pertinent references.

Few suggestions can be given here regarding the use of chemical encyclopedias and dictionaries. Each one has its own peculiarities. All are alphabetically arranged, but each groups topics under major headings, making consultation of the index necessary and the proper selection of search headings mandatory. In Thorpe's "Dictionary of Applied Chemistry", 4th ed., (Vol. I-VIII, A-Oils available), for example, there is a good discussion of the weathering of paint to be found under "Ageing of Paint".

Ullmann's "Enzyklopadia der Technischen Chemie" is valuable for its references to patents, and has a supplement, edited by Siegel, entitled "Verfahren der Anorganisch-chemischen Industrie. . . ." Other useful compendia are:

Dunston, *et al*, "The Science of Petroleum"

Engler and Hofer, "Das Erdöl"

Kirk-Othmer, "Encyclopedia of Chemical Technology"

Tschirch, "Die Harze"

Thus far practically nothing has been said about patent searches. Specifications and claims are written to meet the requirements of the law and their meaning is defined by court decisions which are often more baffling than the claims themselves. Since patents are obtained to shield the exploiter from competition, they are phrased to give, at one and the

same time, maximum protection for minimum disclosure. They generously tell what to use to secure a certain result, but the know-how and economics are lacking.

In spite of all the objections to and accusations leveled at patent systems in recent years, it cannot be denied that the patent literature is vitally important to research. In the first place advances are recorded there long before journal articles appear with accounts of the same results. This fact is well recognized in industrial research laboratories, but too frequently ignored in the academic world "interested in fundamental principles, not their application".

Guides to patent information are provided by each nation offering protection to inventors. Summaries of patents granted are furnished at frequent (weekly) intervals by publication in an official gazette. A complete copy of any patent can usually be obtained upon application to the proper office. Further assistance is offered by detailed systems of classification, utility being the primary criterion. Indexes of various types (i.e., patentee, assignee, subject) are also available, those of the German Patent Office probably being the most varied and complete. Finally, there are many unofficial patent surveys covering restricted fields, such as dyes, aniline, and cellulose. Many such surveys have been published as monographs in Germany. Space limitations forbid the inclusion of an extensive list of patent finding aids. The following indicate what can be expected, details varying from country to country, of course:

Official

United States

Official Gazette, published weekly on Tuesday since 1872. Patents are arranged in numerical order with an index by classification according to use at the end of each issue. Weekly and annual patentee and subject indexes are furnished. Entries for the latter are frequently limited in number to one and are chosen on the basis of utility (e.g., the original stainless-steel patent was put under cutlery, the use advocated in the patent).

Manual of Classification. A compilation of the schedules. Revisions are published frequently in the *Gazette* so that the *Manual* is a looseleaf affair which makes substitution easy.

Patent Specifications and Claims. These may be obtained for a small fee from the U. S. Patent Office, Washington, D. C. Copies of many foreign patents can also be furnished.

File Wrapper. This is the Patent Office record of the history of a patent application. Upon proper petition it may be consulted at the Patent Office after final action has been taken. All applications and correspondence are kept secret until the patent has been granted.

Unofficial

United States

Chemical Abstracts publishes summaries of all chemically important patents issued in all countries. They are indexed in the same manner as other abstracts.

In addition there is an annual patent number index, 1912-14, 1935-19—, and a decennial cumulation for 1937-46.

Special Libraries Association, Patent Index to Chemical Abstracts, 1907-1937.

This fills the gap indicated in the preceding entry.

Abstracts of Chemical Patents Vested in the Alien Property Custodian.

Worden, E. C., *Chemical Patents Index*.

Hohenhoff, E., "Bibliography of Journals, Books, and Compilations which List Abstract Patents."

U. O. P. annual cracking reviews in *J. Inst. Petroleum Technologists*.

A few of the outstanding surveys covering special fields and published in other countries cannot be ignored:

Friedlander, "Fortschritte der Teerfarbenfabrikation. . . ."

Schultz, "Farbstofftabellen."

Rowe, "Colour Index."

Houben, "Fortschritte der Heilstoffchemie."

Bräuer and D'Ans, "Fortschritte in der anorganisch-chemischen Industrie."

Faust, O., "Celluloseverbindungen." 2 vol., 1935.

Gmelin, "Handbuch der anorganischen Chemie." Special patent supplements on iron, aluminum, magnesium, and the platinum metals.

Patent searches are usually initiated to determine the prior art, validity, or infringement in a particular case. Novelty, as set forth in the statutes and interpreted by the courts, is frequently an essential feature, a single prior public use or published description constituting a bar to legal title. This indicates the need for a rapid, preliminary survey of all the obvious sources: highly specialized abstract journals (e.g., Mark and Proskauer's *Resins, Rubbers, Plastics*); a pertinent monograph, bibliography, or handbook (*Chemical Abstracts*); the classified patents in relevant divisions; and specialized journals of the field. If nothing is found in a few hours or days, the problem should be reconsidered from the standpoints of importance, scope, volume of literature requiring examination, allotment of time, and amount of money to be expended. When a thorough search is authorized, the preliminary survey should be reexamined for search words and a definite plan of attack. Indexes quickly give way to a page-by-page examination of likely sections in the abstract journals; however, certain peculiarities do exist: *Chemisches Zentralblatt* published no abstracts on applied chemistry before 1919, when it took over the activities of the "Zeitschrift für Angewandte Chemie," and *Chemical Abstracts* arbitrarily assigns abstracts to the various sections (e.g., puts material of a general analytical nature into Section 7 and that which is more specialized into Section 23 [analysis of paper] or Section 28 [analysis of sugar], etc., as the context may warrant). In other words, the old basic rule applies, search first the most specific, then the more and more general—the box, the shelf, the cupboard, the room, the next room, the floor above, and finally the whole house.

One thing should always be remembered: The abstract journals are guides to original sources, and may be accepted as substitutes only when the originals cannot be examined. In such cases at least two, preferably more, abstracts of the same article or patent should be studied and any conclusions definitely marked tentative.

After the survey of abstract journals has been finished, the references are grouped by journals and the articles examined in chronological order. If necessary, a page-by-page scrutiny of special periodicals is undertaken. Trade newspapers, trade journals, house organs, magazines like *Fortune*, *Endeavour*, and the *Scientific Monthly*, which are not covered by the abstract journals, may be worth inspection. Publications of special service organizations, the *Engineering Index* and bulletins of the U. S. Office of Technical Services, cannot be ignored. Finally, general browsing may yield surprising results.

Some years ago, after many months of arduous and futile searching, a certain investigator was sent to Europe for a much needed rest by his company, deeply involved in a patent suit. While in Vienna he chanced to enter a library and, idly glancing through the *Sitzungsberichte* of a small scientific society, came upon the after-dinner speech of a well-known professor of chemistry. The address contained a complete disclosure of the subject matter in litigation and antedated the patent by several years. A long search and pleasant vacation were terminated simultaneously.

The U. S. Government Printing Office is the largest printing establishment in the world and government publications—federal, state, and municipal—are probably the most numerous; nevertheless, entirely satisfactory and up-to-date finding aids are simply nonexistent. An excellent general survey is to be found in an article by Florence Harden in *J. Chem. Education*, **21**, 326–32 (1944). In addition there are the various lists of publications issued by the Superintendent of Documents, the Department of Agriculture, and the Bureau of Mines. The *Bureau of Standards Journal of Research* and the *Experiment Station Record* are also useful. If one can get his name on the mailing lists of bureaus issuing bulletins of particular interest, he will be kept informed regarding their release; but to ascertain which group of all those “in government” is doing research work in a specific field is not easy. The fault does not seem to lie in any desire for obscurity on the part of a unit, but rather in the complexity of the entire government organization.

One more field of activity remains—searches to reveal the economic aspects of any venture. These range all the way from the selection of a group of products, plant location, construction, and equipment to sales possibilities and competition, costs, manufacturing methods, transportation channels, and financing problems. Two books have been published that

deal with these topics: Tyler, "Chemical Engineering Economics" and Hempel, "Economics of Chemical Industries". A series of nineteen articles on sources of economic information was published by *Chemical Industries* during 1945-7. These articles were subsequently reproduced as a pamphlet, entitled *Information Sources for Chemical Market Research*. The two journals, *Industrial and Engineering Chemistry* and *Chemical Engineering*, contain many articles on American activities, problems, and progress. There are also journals, such as *Modern Plastics*, for individual industries. European countries have similar publications, those of Germany being the most numerous down to about 1942. Many then ceased publication and have not been revived since the close of the war.

All laboratory investigations demand a complete and accurate record of each step in the work. Where legal evidence may be involved, a bound notebook, entries in ink, dates, and witnesses are usually indicated. The same procedure in library studies is well worth the effort. Complete coverage, accuracy of references, timesaving, lack of confusion due to interruptions, and continuity in case more than one searcher is involved are some of the advantages of good records. Several methods have been suggested. One in greater favor involves the use of a bound notebook and punch cards.

General surveys, reviews, and bibliographies are first sought as already indicated. If unsuccessful, the indexes of abstract journals are searched from the most recent back to the last cumulated, then the cumulated as

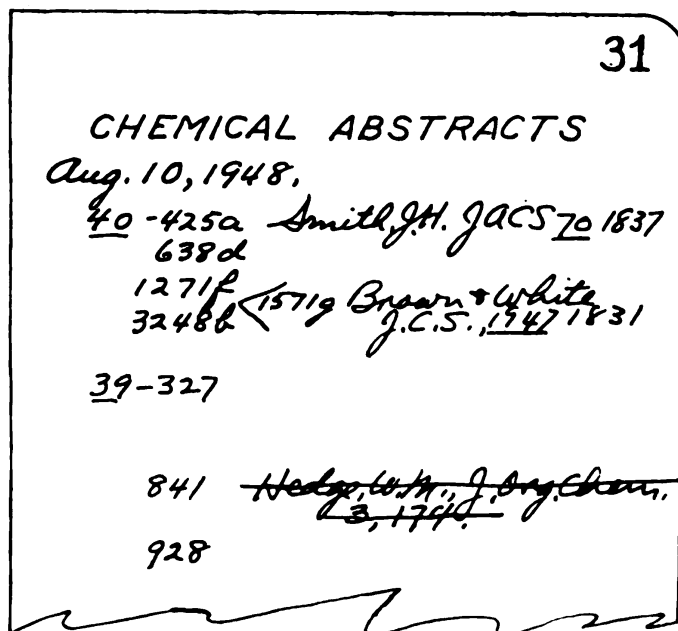


FIGURE 26-1. Sample listing of references.

far back as available or desired. Text references are entered in the notebook, the page numbers for each volume being arranged in ascending order (as shown in Figure 26-1).

After the indexes have been searched, the abstracts are read. When they seem at all promising, references to the original articles are recorded. Next the articles themselves are scrutinized. If irrelevant, the references are crossed out in the notebook. As soon as pertinent data are found, the reference with a statement of its import is put on a card. When the searching is finished, the cards may be arranged in any desired order, a summarizing preface attached, and the whole given to a typist for the preparation of a sufficient number of copies, one of which is kept in the library file for future consultation or expansion. The cards may be kept together or distributed in the subject catalog.

If the searcher and the laboratorian are one and the same person, the preceding steps may be followed down to the examination of the original literature. A decision should then be made regarding the importance of the article. If essential data are given, a reprint, photostat, or filmstat copy should be ordered and a punch card prepared carrying the notation "In file". Time should not be wasted copying extensive details of operations or tables of figures. Punch cards may be made for short notations. These should be so complete, however, that it will be unnecessary subsequently to refer to the original article. During the search an author, formula, or patent-number arrangement of the punch cards and reprint file is preferred to avoid duplication of references as different abstract journals, reviews, and articles are checked one against the other to insure complete coverage. Half to two-thirds of the literature will be quickly and easily gathered. The remainder tests the skill of a good searcher who must finally desist, knowing that no bibliography is ever absolutely complete and hoping that the last four or five per cent contains nothing that will affect the course of the investigation or the final results.*

* For a good bibliography see the excellent article "Library Techniques in Searching" by D. F. Brown in "Advances in Chemistry Series, Vol. 4, Searching the Chemical Literature," pp. 146-57, Washington, D. C., American Chemical Society, 1951. Revision in progress.

Chapter 27

TRANSCRIPTION PROBLEMS IN PREPARING AND USING PUNCHED-CARD FILES

C. D. GULL*

National Academy of Sciences-National Research Council, Washington, D. C.

Introduction

Two aspects have been stressed so far in this book about the application of punched card devices and other mechanical selectors in the organization and control of scientific and technical information, namely, methods of recording information as notches or punches in the cards, and methods of selecting and arranging all the cards on a given subject by means of the notches or punches. These are intellectual and mechanical aspects of the storage and retrieval problems associated with punched cards.

The transcription of information is a third important aspect to be discussed in this chapter. There are two parts to transcription: (1) copying onto punched cards for the creation of records and (2) reproduction from punched-card files as a service to users. Transcription problems exist because the technology of 1957 is not sufficiently advanced so that one method can be used to create a record and put the information on cards from oral, graphic, mechanical and electronic sources, and then to reproduce that record in any quantity whenever necessary. Because no one method is available, librarians and information specialists must select from a variety of methods and combine them to meet their needs.

The chief expense involved in establishing a punched card file is not the cost of the cards or the cost of putting the information on the cards and of reproducing it. The chief expense lies in obtaining the information to go on the cards, involving as it does literature searching and coding which must be done by technically trained persons. If cards can be duplicated and reproduced readily when necessary, the original cost is distributed over all the sets of cards made; large savings in scientific manpower can then be effected, since it is no longer necessary for each individual user to do his own searching, abstracting and coding.

The principal objective of this chapter is to describe the various methods for recording and reproducing the graphical text of punched cards and to

* In 1958, Mr. Gull joined the Computer Department of the General Electric Co.

relate the methods to the types, capabilities and limitations of punched cards. The novelty of punched cards may obscure for inexperienced users the fact that most of the information contained on them is written there in some fashion and that the notches and holes are only guides to the written information. Punched tabulating cards are an exception to this generalization whenever the written information is identical with the punched information, for it was transcribed or interpreted directly from the holes.

Punched card files require more careful attention to transcription problems than customarily devoted to most conventional card catalogs which are made of multiple copies of the records created to represent the books, pamphlets, reports, etc., of a library or information collection. Conventional practice rarely anticipates the legitimate need of catalog users to take away copies of the records they select. Some information services store copies of the catalog cards to be given away or sold, but most users are expected to copy by hand or resort to photography; or more likely, to depart with less information than they can profitably use.

Multiple copies of catalog cards are created to be filed in arbitrary arrangements, usually alphabetic or numerical, under entries for authors, titles, subjects, accession number, etc.; such cards are often locked in place by rods to make sure that careless users do not misplace or make off with them. Requests for permission to photograph cards are unwelcome, because the cards must be unlocked, removed and later carefully refiled in their proper places.

In contrast, punched cards afford the opportunity to create files of reduced size; only one card is required for each item in the collection, yet by means of notches or punches, equal or greater access is provided to documents under entries for authors, subjects, etc. In many instances it is not necessary for the cards to be filed in any fixed order. Whenever users select and remove punched cards from a file, however, they deprive others not only of the cards removed but also of access to all corresponding documents, for there are usually no duplicates of the removed cards which can be consulted in another part of the file. This situation requires that the librarian or director be prepared to reproduce parts of the punched card files rapidly under maximum demands for service, and preferably at minimum expense. Users may ask for reproduction on sheets of paper, or they may ask for full or partial, punched or unpunched, card files for use in other locations.

Transcription Methods

The methods by which transcription can be accomplished for punched card files are derived from the graphic arts, photography, electronics, telegraphy, and machine accounting.

The graphic arts provide a choice or combination of handwriting, type-writing, letterpress printing (hand and machine set type, Multigraph), spirit duplication (hectography), stencil duplication (mimeography), direct offset lithography (Multilith), and embossed plate printing (Addressograph). In these methods the characters are created one at a time by hand.

Photography can be used to copy one or more pages of existing records onto negative film and positive photopaper by conventional wet processes, and there are variations in the blueprinting, Ozalid, diazo and microphotographic (microfilm and Microcard) processes. Photography is also used in printing by photo-offset lithography.

Xerography is a dry, electrostatic photocopying process for producing single copies and for printing by XeroX-offset lithography.

Thermofax is a dry, infrared photocopying process for single copies. It does not satisfactorily copy aniline dyes which are used in writing inks, some ball point pens, spirit duplication, and in some typewriter ribbons. Storage near excessive heat will darken the print, sometimes completely obliterating the information copied.

Facsimile is an electronic copying process. While facsimile can be used to produce single copies and to cut stencils for use within the library, its greatest usefulness lies in its ability to transmit messages over considerable distances on wire or radio circuits. Television, Telefax, Wirephoto, Radiophoto, Timesfax, and Ultrafax are examples of facsimile applications.

Photography, xerography, Thermofax, and facsimile can be used to create original records on punched cards (or records to be affixed to punched cards) as well as to copy the texts from punched cards.

For about half a century, machine accounting with International Business Machines (IBM) and Remington Rand equipment has provided automatic mechanical transcription controlled by the holes punched in tabulating cards, and for about fifteen years it has been possible to punch tapes from cards, and vice versa, for the control of teletypewriters on local and long distance circuits. It is the widespread availability of equipment and automatic mechanical transcription which have made machine-sorted installations so attractive to persons who are responsible for operating large-scale information services. Automatic mechanical transcription can be accomplished a line at a time from the holes punched in tabulating cards but only a character at a time from the teletype tapes.

In many places machine accounting operations are being converted to the use of electronic digital computers, in which the punched holes are frequently replaced by magnetized patterns on metal and plastic tapes. But the conversion to computers does not eliminate the transcription problems, which are known as the input and output operations of computers. It is extremely likely that the various solutions which will be de-

veloped for the input and output needs of computers will be adaptable to the transcription problems of punched card files.

Thus it can be seen that there is considerable choice of transcription methods; however, they vary widely in suitability, ease of use, and expense in relation to different situations and different punched-card applications.

Hand-Sorted Cards

In the discussion of transcription problems of hand-sorted cards, much of the information applies also to the unpunched portions of tabulating cards.

In duplicating cards the notches have to be made in an operation separate from printing the text. The conventional hand punch is used when the number of duplicates is small. Gang punches, already described on page 34ff., are available for duplicating the notching of many copies of a single card at a single stroke.

The conventional graphic arts duplicating methods already mentioned are so well known as to need little discussion here. Some of the most commonly used methods are described in recent articles on catalog card reproduction¹. Most of these methods, including the preparation of multiple carbon copies on the typewriter, are satisfactory when the number of cards needed is known. Since in many cases the ultimate number cannot be estimated, and since it may not be desirable to store copies against future need, it is a definite advantage to be able to duplicate a card in one operation when needed, rather than to repeat the manual operations involved in preparing the original card. This ability is particularly advantageous when an entire file of cards is needed. Hectograph spirit paper masters and Addressograph plates can be stored for future use and are convenient to use. Stencils and Multilith masters can also be stored, but are somewhat less convenient to use.

In the following paragraphs, emphasis is placed on the methods which can be used to prepare and duplicate cards by copying large areas of text rather than by creating one character after another manually. Since these methods are best adapted to preparing single copies, it is advantageous to combine them with stencil reproduction or offset lithography whenever it is necessary to obtain multiple copies².

There are two types of copying equipment, directly reflecting the bound or unbound state of the material to be copied. In the rotary type of equip-

¹ Hensel, Evelyn, *et al.*, "Catalog Card Reproduction," *J. Cataloging and Classification*, 12, 209-220, Oct. 1956.

² Lewis, Chester M., and Offenhauser, William H., "Microrecording; Industrial and Library Applications." New York, Interscience Publishers, Inc., 1956.

Doss, Milburn Price, ed. "Information Processing Equipment," New York, Reinhold Publishing Corp., 1955.

ment, single cards or single sheets of paper are transported around cylindrical surfaces while being copied. In the flat-bed type the pages of bound materials are pressed against flat glass surfaces. Cards can be copied by either type, but books can be copied only on the flat-bed type. If the flat-bed type is equipped with a lens system and is adjustable for focal length, copies can be reduced in size or enlarged, but most of the rotary types are only able to produce copies which are the same size as the original.

There is a definite advantage in being able to copy an abstract from, for example, *Chemical Abstracts*, directly onto a hand-sorted card. Such copying avoids manual transcription with subsequent proofreading and correction; it is particularly desirable when transcribing chemical equations and organic formulas. In addition to being used to prepare cards by photo-offset lithography, a camera can be used to produce a microfilm negative. The Filmsort Division, Dexter Folder Co., Pearl River, New York, offers a service in which four frames of 35-mm microfilm can be mounted in a marginal punched card, to be read with their own reader. For a few years after World War II single frames were also mounted for use in IBM punched cards. Many of the microfilm readers can also be used as enlargers to prepare positive prints from negatives mounted on cards or from negative film strips stored in pockets of cards. RCA demonstrated an Electrofax enlarger printer designed for use with the Filmsort cards^{2a}.

Positive prints of microfilm strips can also be produced as Microcards of various sizes to be mounted on marginal punched cards or in long rolls as Microtape. Microtape is cut to the proper length for the card and mounted with its own adhesive, just as Scotch cellophane tape is used. The adhesive has a guaranteed life of 25 years. Microcards and Microtape are available from the Microcard Corp., West Salem, Wisconsin. They add weight to marginal punched cards and reduce the failure of cards to drop from the needles.

Xerography offers the singular ability to produce one electrostatic copy on any kind of untreated card stock used for marginal punched cards; that is, the paper need not be light sensitive. The powder image is fixed on the paper by heating for about 30 seconds. The Haloid Company of Rochester, New York, produces both contact printers and cameras for this process. If many copies are needed, the image is transferred from the exposed selenium plate to a paper Multilith plate rather than to a punched card, and the Multilith plate is used for printing cards by offset.

An RCA flat-bed facsimile scanner was used experimentally at the Library of Congress during 1954. The image was sent about twelve miles to the National Institutes of Health Library where it was received by the

^{2a} Anon., "Electrofax Dry-photographic Enlarger," *J. Franklin Inst.* 261, 584-5, May 1956.

electrolytic process. The received images can be mounted on marginal punched cards by using cements, mounting tissue, or a laminating process. An RCA facsimile hand scanner was under experimental development in 1949 but was not yet commercially available in 1957. The scanner could be passed rapidly and selectively over the words to be copied, permitting rearrangement of the text to suit the user's needs; it also made it easy to copy from bound volumes³.

Reflex photocopying with flat-bed equipment is relatively common in such machines as Apeco, Contoura, Copease, Cormac, and Verifax; it involves the transfer of a light-sensitive dye from the negative sheet to the positive sheet to form the image of the original. One negative is commonly used in preparing one positive with these machines, but Verifax papers permit as many as ten copies from one negative. Contoura is available from Frederick G. Ludwig, Inc., Deep River, Connecticut, and Verifax from Eastman Kodak Co., Rochester, New York. Reflex photocopying and Thermofax produce copies of abstracts and other texts on paper which must be mounted on punched card stock. Recently, however, a method has been described of transferring an image from a Verifax negative directly onto punched cards which are made from a stock which was found to have the absorption and other suitable properties without being treated or sensitized in any way^{4a}.

The Ozalid photographic process, one of the rotary types, can be used to prepare translucent and opaque hand-sorted cards. The original text is typed on a translucent Ozalid master sheet or card. Positive copies of cards are produced by repetitively transporting the master and a sensitized card together over a cylindrical surface and exposing them first to light and then to ammonia vapor for development. If the punched card file is maintained on translucent cards, selected information can be transferred to Ozalid positive paper for use as a search report. The M. W. Kellogg Company, Jersey City, New Jersey, and the then Central Air Documents Office at Wright-Patterson Air Force Base, Dayton, Ohio, have made use of Ozalid cards. Literature describing the Ozalid process may be obtained from the General Aniline and Film Corp., Johnson City, New York, and both translucent and sensitized Keysort cards are available from the Royal-McBee Company, Athens, Ohio.

A similar method is the semidry, diazo photographic process developed by Dr. Van der Grinten of Venlo, the Netherlands. It is available in the United States as the Copyflex process of the Charles Bruning Co., Inc., of

³ Taube, Mortimer. "New Tools for the Control and Use of Research Materials." *Proc. Am. Phil. Soc.*, **93**, 248-252 (1949).

^{4a} Passer, Moses (University of Minnesota, Duluth), "A New Photocopying Process for Punched Cards," *J. Chem. Educ.*, **33**, 581-3, Nov. 1956.

Chicago. In this method, the abstract is copied by reflex photography through a reflex film intermediate onto the translucent master card. The master card can then be used to make duplicate cards or to transcribe the information onto paper. Punched cards sensitized with the proper diazo dyes are available from the Royal-McBee Company. De Gorter states that the diazo process is being used at Imperial Chemical Industries Limited to print bibliographies directly from punched cards. The cards are arranged and six to eight cards are printed at one time on diazo-copying paper⁴.

After typing the original text onto a translucent sheet or card, an abstract or any printed text can be copied onto the same sheet or card by using xerography. Positive copies of the ingeniously combined text can then be made as desired by the Ozalid and Copyflex processes, providing additional cards or the sheets of a report.

Another method of copying, e.g., abstracts from *Chemical Abstracts*, is by use of the Polaroid Land camera, which produces positive prints soon after the exposure has been made. Special supports and apparatus may be used to facilitate large-scale operation.

Three methods employed in preparing bibliographies and search reports from punched card files are worthy of description here. The oldest of these methods is used extensively by the Library of Congress in Washington in preparing its various cumulative catalogs which list books, pamphlets, maps, atlases, sound recordings, motion pictures, etc. The unpunched cards, previously arranged by hand into author and subject sequences, are arranged in one to four columns and fastened onto cardboard sheets with Scotch masking tape, while the sheets are held in a card aligning machine. The sheets are then photographed and the catalogs are printed by photo-offset lithography. After the tape is removed, the cards are returned to file boxes until they are needed for the next cumulated issue of a catalog. One person can prepare 100 pages a day for the camera, with an average⁵ of 40 entries per page. This method was also used at the Library of Congress to prepare the abstract pages as well as the indexes of the "Technical Information Pilot"⁶. Whenever this method is applied to hand-sorted cards, it is necessary to arrange the punching and text areas on the cards so that the cards can be overlapped in columns without leaving too much blank space in the finished reports and bibliographies.

Facsimile is used in the second of these methods. After the cards are

⁴ De Gorter, B., "The Principles and Possibilities of Diazo-Copying Processes," *J. Documentation*, 5, 1-11 (1949).

⁵ Gull, C. D. "The Cumulative Catalog Technique at the Library of Congress." *Am. Doc.*, 2, 131-141, August 1951.

⁶ Taube, M., "The Planning and Preparation of the *Technical Information Pilot* and Its Cumulative Index," *College and Research Libraries*, 9, 102-206 (July 1948).

selected and arranged in the conventional manner, they are placed in a facsimile scanner. The product of the reproducer, somewhat reduced in legibility, is a long roll of paper containing the references and abstracts in the order in which they were fed into the scanner. The paper is cut to page size and the search report is assembled in booklet form. This method is used by some organizations with machine-sorted cards. It can be used to distribute information very rapidly by wire or radio from a central information service.

In recent years the overlapping, manual, cumulative catalog technique has been mechanized by combining machines to provide justified proportional typewriter composition, machine sorting and searching of cards, and automatic feeding exposure of cards in automatic cameras. One to three lines of text are copied from each card onto long rolls of photographic film in the automatic cameras. The developed negative rolls are cut and mounted in the desired number of columns for the selected page size, and the resulting catalogs are printed by photo-offset.

The Friden Calculating Company of San Leandro, California, offers the following combination of equipment for this third method. The typist sets copy on the Justowriter Recorder Model "L" in one, two or three line increments. If satisfied with the proof visible in the typewriter, the typist transfers the text to a roll of card stock in the Justowriter Reproducer through a punched paper tape. Justification and error correction are accomplished through the tape. An automatic line finder insures proper placement of the increments on the roll of paper. The finished roll of card stock is then die cut to IBM tabulating card size on a cutter available from the Standard Register Co., Dayton, Ohio. The resulting cards are then coded in column 52 to indicate the number of lines of text, and with additional punching so that the cards can be used in sorters, collators, etc. Whenever a catalog or list is needed, the cards are automatically fed into the Kodak Listomatic camera where they are photographed at the rate of 230 cards per minute, at the same size as the original or reduced as much as 50 per cent. The cards are returned to storage, for addition or deletion of cards by hand or mechanically with the sorter and collator.

The Ralph G. Coxhead Corp., Newark, New Jersey, offers a similar combination through the use of its Varityper composing machine, hand-sorted or machine-sorted punched cards, and an automatic camera. While this combination is limited to one line per card, it permits the use of any kind of cards within a convenient size range.

Machine-Sorted Cards

Punched information in machine-sorted cards can be duplicated very inexpensively if the proper type of punch is obtained. The "IBM" alpha-

betical printing punch has a duplicating feature which allows the entire card to be reproduced in 8 seconds. The "Remington Rand" alphabetical punch will duplicate cards at the rate of 90 per minute. With this equipment the only additional expense for duplicating is for the extra cards and the operator's time. More elaborate equipment, such as the "IBM" Card Reproducing Punch or the "Remington Rand" Multicontrol Reproducing Punch, is available for reproducing decks of cards or for gang punching at the rate of about 100 cards per minute. These latter machines are capable of other operations. For example, the Multicontrol Reproducing Punch is also a collator which may be used instead of the sorter as a card selector.

One interesting development provides an economical, efficient means for conveying information punched in "IBM" cards from one location to another without the necessity for sending the actual cards⁷. A card-controlled tape punch is used to prepare from the cards a paper tape containing all the information to be transmitted. One reel of tape 8 inches in diameter and less than an inch wide contains all the information punched in 1500 "IBM" cards. This tape is mailed to the subscriber. There the tape-controlled card punch is used to interpret the tape automatically and punch the information into "IBM" cards. It is also possible with these same machines to send the information between two points by commercial wire service, an operation which is performed automatically. The significance of this device for speeding the transmission of information from a central bureau is evident. This equipment has been used at the Wright-Patterson Air Force Base, Dayton, Ohio. This punched tape may also be used with the "Cardatype" machine mentioned later.

The transcription of the information from a card to a piece of paper depends on whether the information is punched in or recorded on the face of the card. Both types of systems are in use. Major emphasis will be placed on the case where the information is punched in the card since what has been said about hand-sorted cards applied to machine-sorted cards when the information is typed on them.

Two types of demands may be made of the cards once they have been selected and arranged in useful order: The person who made the request may be satisfied to glance at the information on the cards and make notes, or he may ask for a printed report.

The first demand specifies that the information be placed on the card in readable form since it is difficult to gain any facility in interpreting directly from the punched holes. The printing punch ("IBM") or interpreter

⁷ "Proposed Operating Procedures for Initial Installation of Types 041 and 057 Machines as Applied to Freight Train Operation on the New York, New Haven and Hartford Railroad." Electric Accounting Machine Division, International Business Machines Corp., New York, N. Y. 1945.

("IBM" or "Remington Rand") will print on the card the letters and/or digits punched in the card. Additional information may be typed on the cards, but this method cannot be used to activate machines. It is possible, by using several cards as a bibliographical unit, to read both the reference and an abstract from the interpreted cards⁸.

In general, it is desired to record the results of a literature search as a written bibliography or search report. It is true that the information appearing on the interpreted cards may be copied by a typist, but this is inefficient if the operation is performed often enough so that the cost of performing the operation mechanically can be amortized readily. Thus far there are two machines for doing this—the tabulator and the card-operated typewriter. The tabulator is a device in which type bars or type wheels, activated by the holes punched in the card, are used to print on a sheet of paper the letters or digits punched in the card. Various models of tabulators are available from "IBM" and "Remington Rand". The "IBM" type 405 tabulator is limited to printing 43 alphabetical characters in a single line, and 80 arabic numerals. The type 403 tabulator prints three lines from a single punched card, so that alphabetical information can be obtained from any column of the "IBM" cards. The largest "Remington Rand" tabulator can print as many as 100 alphabetic symbols in one line. The typography is very limited since only 26 letters of the alphabet (in capitals), 10 numerals, and 3 punctuation characters can be punched into cards and used for printing. The printing is done at the rate of about 100 lines a minute. Examples of the appearance of such printing may be found in Reference 8. Crandall and Brown⁹ at the Socony-Vacuum Laboratories have tried to use this method to print the *Index of Current Technical Articles* directly from "IBM" cards. The authors state that the adoption of the method is delayed on the problem of typography. The technical staff of their organization is reluctant to accept a publication printed entirely in upper case with little or no punctuation. "IBM" also offers an electric accounting machine, type 407, which overcomes some of the disadvantages mentioned¹⁰. The type bars have been replaced by type wheels. Each of these type wheels is equipped to print the complete alphabet (in capitals), numerals, and 11 special characters. With this machine 120 alphabetical-numerical characters may be printed in one line. There are 10 printing positions to an inch, as in pica typewriter spacing. The speed of operation has been increased to 150 cards a minute.

⁸ Gull, C. D., "A Punched-Card Method for the Bibliography, Abstracting, and Indexing of Chemical Literature," *J. Chem. Ed.*, **23**, 500-507 (1946).

⁹ Crandall, G. S., and Brown, Betty M., "An Information Service Using Both Hand- and Machine-Sorted Cards," *J. Chem. Ed.*, **25**, 195-99, 203 (1948).

¹⁰ "New 'IBM' Accounting Machine," International Business Machines Corp., New York, N. Y.; "IBM' Accounting Machine, Type 407," *Preliminary Manual of Information*, New York, International Business Machines Corp. (1949).

Crandall and Brown in the same article describe a special-purpose file, in which information about petroleum additives contained in the patent literature is coded on "IBM" cards. The tabulation of this set of cards by the various criteria coded shows the rapid mechanical provision of various types of listings from the same cards. In this particular case lists were compiled chronologically by patent number, assignee, product, patent class, function, and chemical composition of additive. It would be very time-consuming to make such listings by conventional methods. Listings of this type make it possible for new relations to be discovered between structure and additive properties. By the use of the multiple selector and tabulator it is possible to print lists of references having a predetermined combination of codes, such as all metallo-organic compounds used as anti-knock agents in petroleum fuels.

The then Central Air Documents Office used this method for compiling bibliographies, a foreign-language dictionary, desk catalogs of captured German documents, etc.¹¹. The method is ideally suited for compiling indexes, subject-heading lists or mathematical tables where typography is not so likely to be criticized. The value for this purpose can be understood when it is realized that the cards may be arranged and interfiled mechanically so that a cumulated index or list of headings may be printed at any time. If the cards are verified after punching, proofreading and editing are unnecessary. The index to the "Classification of Patents," published by the U. S. Patent Office, is produced by the following: The index entries are punched on cards, and the cards are then arranged and tabulated. The tabulation is then photographed and the index published by offset lithography. The process is described in detail by Cochran¹². Other examples of indexes compiled in this way are to be found in *Nuclear Science Abstracts*, published by the Technical Information Branch, Atomic Energy Commission, Oak Ridge, Tennessee¹³.

A second device which may be used in place of the tabulator for printing the information punched in the cards is the "IBM" "Cardatype"¹⁴. An earlier model, which was loaned to the Library of Congress by the International Business Machines Corporation for experimentation and which is still in use there, consists of two units: the "IBM" electric typewriter and, connected by an electric cable, the "IBM" verifier, modified to read the holes in a punched card and now called the "reader". The latter feeds one card at a time off the bottom of a pack of cards, placed face down and

¹¹ Lutz, A. W., *Mechanized Processing of Air Technical Documents*, paper read at A.C.S. Convention, Atlantic City (Apr. 16, 1947).

¹² Cochran, S. W., "Recent Progress in Patent Classification," *Ind. Eng. Chem.*, **40**, 731-33 (1948).

¹³ *Nuclear Science Abstracts*, **1**, 12 (Dec. 30, 1948).

¹⁴ "New IBM 'Cardatype,'" International Business Machines Corp., New York, N. Y. (August, 1949).

arranged in the normal order of filing, that is, 1 to 10, etc., or A to Z. Unlike the tabulator, which reads all 80 columns at once and prints 80 characters simultaneously from one card, the reader feeds the card lengthwise one column at a time. Hence the typewriter, which can print only one character at a time, is restricted to a speed of only 600 characters per minute compared to 6400 characters per minute for the tabulator.

Various combinations of holes in the cards are used to control all the motions of the standard typewriter (i.e., the 84 characters, shift, shift release, carriage return (which includes line spacing), columnar tabulation, and back spacing). Additional combinations control certain operations, such as causing the reader to skip portions of the cards and eject them. In the commercial model, a program tape is used to control the operation of the machine, which will stop at predetermined points to allow for the manual typing of variable information. The electrical impulses are controlled through electromagnetic relays and the customary interchangeable plugboards which are a feature of most "IBM" machines.

In the "Cardatype," a tape-punching unit is also provided for automatic recording of selected portions of the typing because, in most applications, the information will be used in subsequent punched-card operations at some point other than that at which the "Cardatype" is located. The transmittal tape can then be mailed conveniently and with small expense and the information transcribed back into punched holes by the tape-controlled card punch, mentioned previously⁷.

The typewriter can be used, of course, for any operation normally performed manually on any typewriter. Typing can be done upon sheets, cards, stencils, and also on Multilith mats, and various ribbons can be used. The usual number of carbon copies may be made. The automatic typing is done at the rate of 10 characters a second. If the proper characters are placed on the typebasket, it appears that most alphabetic languages and many mathematical symbols can be accommodated on the card-operated typewriter. An illustration of this kind of machine, and a description of its use in printing the "American Air Almanac," may be found in an article by Eckert and Haupt¹⁵. The article discusses the advantages of this method of printing tables over conventional methods

A punched-card system in which the information is typed on the card is described by Peakes¹⁶. The holes in the cards are used only for selection of the desired cards. Such a system avoids the use of the tabulator, which is the most expensive piece of equipment. It would appear that fairly extensive use would be required to justify renting the tabulator. Peakes

¹⁵Eckert, W. J., and Haupt, R. F., "The Printing of Mathematical Tables," *Mathematical Tables and Other Aids to Computation*, 2, 197-202 (1947).

¹⁶Peakes, G. L., "Report Indexing by Punched Cards," *J. Chem. Ed.*, 26, 139-146 (1949).

states that an abstract of about 125 words may be typed on the card. A typewriter with $\frac{1}{4}$ -inch line spacing is needed. This can be had by double spacing in a micro-typewriter, a standard but not common design offered by several manufacturers. The methods that could be used to duplicate this type of card would be a combination of the methods used for hand-sorted and machine-sorted cards. The methods required for transcription to a sheet of paper would be the same as for hand-sorted cards. Facsimile machines have been used for transcribing information from this type of card.

Other Devices Employing Punched Cards

The microfilm rapid selector^{17, 18} employs a punched card mask in searching for entries and abstracts stored on 35-mm microfilm. As the film is transported rapidly past photocells, the mask is used to search the digital index pattern of black and white dots on the film which is used to identify the subject matter. When a pattern and the mask fully match or complement each other, the photocell triggers a high speed flash to photograph the corresponding entry and abstract on unexposed microfilm, which is then developed and printed and enlarged on paper for reading. The mask is punched on IBM equipment. The rapid selector was still under experimental development in early 1957.

Electronic digital computers¹⁹ are used in ever-increasing variety and complexity; and machine-sorted punched cards and punched tapes are used in many computers to increase their storage capacity and to reduce the time needed to feed information into computers and to read it out of them. As described earlier, cards can be made from tapes and tapes can be made from cards. They are interchangeable in many uses, although tape speeds exceed card speeds. The input transcription problem is confined to initially punching the holes in cards and tape by hand, although attempts are being made to enable computers to read written characters. Cards and tape are used to actuate typewriters and tabulators to transcribe or read out selected information, and the day is fast approaching when computers will provide written displays on electronic picture tubes²⁰.

The great potentialities of computers in information services have been

¹⁷ "Report for the Microfilm Rapid Selector," Engineering Research Associates Inc. Available as PB-97313 from Office of Technical Services, Dept. of Commerce, Washington, D. C.

¹⁸ Wise, Carl S., and Perry, J. W., "Multiple Coding and the Rapid Selector," *Am. Doc.* 1, 76-83 (1950).

¹⁹ Pinkerton, J. M. M. "Recent Developments in Electronic Computers," *Endeavour*, 16, 36-41, January 1957.

²⁰ Storage Tube Named "Memotron" Can Capture and Retain Visual Displays of Transients Without Need for Photography. Developed by Hughes Aircraft Co. Los Angeles 45. *Research and Engineering*, 2, 47, June 1956.

demonstrated in the use of the Univac computer to prepare a concordance to the Revised Standard Version of the Bible in a year or two, compared to the 30 years of effort that went into Strong's "Exhaustive Concordance" of 1894. In this effort accuracy was assured by copying the words of the Bible plus the identifying book, chapter, and verse, first onto magnetic tape and then again onto punched cards. The cards were converted to tapes and all discrepancies uncovered by electronic comparison of the two tapes were then corrected. The Univac was then used to alphabetize the words, eliminate 132 unwanted words, and transfer the wanted words in order to an output tape, from which the Unityper transcribed the 350,000 words and references in 1000 hours for use by the typesetters²¹.

²¹ McCulley, William R. "Univac Compiles a Complete Bible Concordance." *Systems Magazine*, 20, 22-23, March-April 1956.

PART IV

FUTURE POSSIBILITIES

CHAPTER 28

EVALUATION OF MECHANIZED DOCUMENTATION AT THE GMELIN INSTITUT

E. H. ERIC PIETSCH

Gmelin Institut für Anorganische Chemie und Grenzgebiete in der Max-Planck-Gesellschaft, Frankfurt-am-Main, Germany

Introduction

The second edition of this book has required a considerable revision of this chapter. This necessity clearly indicates that the field of documentation has been undergoing rapid development during the intervening years. Our civilization is being profoundly influenced by technology, by automation and by continuous extension of human control over the forces of nature. R. Jungk has followed these developments for many years with penetrating insight and exceptional knowledge¹. Paraphrasing one of his remarks, we might say that, in documentation, the future has already begun. Considerable diversity of experience in developing automatic documentation methods, since publication of the first edition of this book, enables us to make more specific predictions as to future possibilities than could be made a decade ago or even as recently as five years ago.

Recent progress in documentation has not been free of disappointments and there has been considerable disagreement—even conflict. It is possible to point out various reasons for this. With few exceptions, the new science of documentation has attempted to apply for so-called mechanical documentation apparatus and equipment which was already available and which seemed—at least at first glance—to be suitable for providing solutions to the problems of controlling recorded knowledge. However, during recent years, it has become more and more generally understood that, for those operations which are essential to documentation and which in addition are so closely linked to the structural operating procedures of human intelligence and the human spirit, it is necessary to develop radically improved procedures and equipment. Otherwise a satisfactory solution can scarcely be developed for our general problem, which perhaps is best formulated as surmounting the excessive demands made on human memory by the continually increasing amounts of recorded information generated by observation, by experiment, and by investigative study in general.

A further factor of basic importance in this connection has to do with the nature of the "classical" methods of documentation as previously applied. These methods are based on principles of analyzing and ordering the factual and the conceptual contents of recorded knowledge. These principles could be accepted as valid as long as we could successfully apply linear arrangement as the basic principle for accomplishing the selection of needed recorded knowledge. Once we begin the attempt to create an external form of memory, not based on some form of linear arrangement, we encounter problems of a different order of magnitude. In solving these problems, we must arrive at other principles for the analysis and the harmonious correlation of data, observations and facts. For this reason, the creation and the establishment of automatic documentation systems require, as an unavoidable prerequisite, the development of a system for the control of "documentation elements". The new system must apply principles completely different from those that have previously been customary in indexing, classifying and codification even though previously formulated principles have been successfully applied in the past. Here it must be emphasized that we cannot be satisfied with partial solutions for immediate problems. In this connection a particularly difficult situation is encountered. The documentalist, in his professional activity in industry, science and technology, is called upon to provide solutions for immediate problems. In other words, the documentalist is confronted with the requirement to justify his day-to-day efforts. As a consequence, not only is the development of fundamental solutions placed in jeopardy but the documentalist, by developing improvements that are only partially successful, runs the risk of appearing in a false light or—from a somewhat different point of view—there is danger that he will place modern documentation as a whole in a false light. What is essential and necessary in this situation is recognition that documentation must be regarded as an important field of research in its own right. More specifically it must be recognized that considerable financial support will be required for the solution of documentation problems which are of decisive importance for efficient and economic accomplishment of creative work in all fields of intellectual and professional activity. At the same time, it is essential that documentation be accorded the status of an important branch of instruction at our universities and colleges. In this connection, it is particularly important that courses of instruction be worked out which will attract young people to a professional career in documentation. We are already faced by the threat that the development of documentation methods and systems may prove an empty accomplishment because of the dearth of active documentalists to assume the tasks of leadership and administration.¹⁴

Misunderstanding or disregard of these points has led, during recent

years, to various disappointments which should not be overlooked during future development of documentation as a field of investigation. On the other hand, the achievements of recent years have made it clear that we are not exaggerating when we speak of mechanical documentation as an already accomplished fact. Today systems and methods are available which can ease considerably the burdens imposed by our ponderous literature. Furthermore, the trend of future development can also be clearly discerned. It is particularly gratifying to note in this connection that the future of documentation will not be limited to the development of technical aids. Rather, a wide field for intellectual accomplishment has been opened up. Studies on the nature of information, language and its laws, and translation from one language to another with mechanical aids, have already provided glimpses of the broad range of problems whose solution will lead to insights of fundamental importance.

The Document—Definition of its Character¹

As long as oral transmission provided the only means for communicating observations, ideas, concepts, cause and effect relationships or similar results acquired by the human senses or created by thought processes, these various forms of human experience necessarily shared the transitory character of the individual person or of human groups. The results of human experience, in such transitory or unrecorded form, could scarcely attain that measure of permanence and stability that is essential to the building blocks of intellectual achievement. This observation remains true even for exceptional cases, when strictly maintained discipline of the memory transmitted extensive intellectual treasures from generation to generation. Here we might mention the preservation of religious or historical legends and similar traditions as exemplified in heroic poems. A specific, particularly impressive example that one might mention are the great Indian epics, the Ramayana and the Mahabharatam. But once such intellectual or cultural treasures have been fixed, once they have been freed of the limitations of the individual, temporally limited person, then they achieve as large a measure of permanence as is humanly attainable and, at the same time, are transformed in objective character. This act of fixing or recording observations, ideas, concepts, cause and effect relations and the like for the purpose of preserving and removing them from the sphere of the transitory into the realm of intellectual permanence leads to documents. The creation of documents—simultaneously an essential act of both self-expression and of development of human potentialities—began in the earliest stages of human history even though sometimes not recognized in its true nature for long periods of time.

The cave paintings of paleolithic times resulted from human response

to the surroundings on an emotional, sensual (plastic) or magical basis. Nevertheless, this form of representation was not static in character but represented a dynamic interaction between man, on the one hand, and the object being portrayed, on the other hand. For example, the quarry of the hunt is represented as hindered in its flight by a symbolic arrow in the body of the animal or the picture of the animal is shown close to a rocky chasm and so on. In neolithic times, we observe that man entered into a further stage of development of consciousness with a clear distinction between the individual ego and the environment. A very broad range of documentary material from all parts of the world shows how man interpreted his life in this world during the later epochs of the Stone Age and also his relationships with supernatural forces, gods and demons. Here again, drawings and paintings on stone provide us with considerable insight into millenia of prehistoric human development which preceded the development of the human cultures of early historical times. In this connection it is particularly important for us to note that there is a continuing tendency for the graphically portrayed subject matter to become simplified as human development continued through the years. There is a tendency to restrict the subject content of graphic material to essential features and thus to arrive at symbols for various ideas or thoughts. As an example, we might cite the stone drawings from the North African so-called Capsien (Tfongs)^{3, 4}.

The flexibility of the means of expression remains naturally limited as long as one is restricted to a pictorial representation. But even at this stage, there is an easily observable tendency to strive to establish basic ideas, units of thought—a tendency which springs from human analytical power. This tendency has been observed to be well developed at the dawn of history as we understand it, that is to say, at a time when the individual begins to emerge from his previous anonymity. In this connection, it is astounding to note that, in the case of hieroglyphics, neither Herodotus nor Strabo nor Diodorus and not even Herodotus, of whom only the latter exhaustively described the hieroglyphics and all of whom were much closer to the Egyptians than our nineteenth century culture, took notice of the fact that the Egyptians had already taken a decisive step in the development of writing. A long period of confusion on the part of European scholars was finally brought to an end by the bold conception of Jean François Champollion⁵ that the hieroglyphics achieved the decisive analysis of concepts and their expression as words, symbols and corresponding units. In this way the analytical representation of phonetic complexes was made possible as well as the use of individual symbols to create complex statements and documents. The development of the Assyrian Babylonian cuneiform by the Sumarians also accomplished the same break through

from the pictorial representation to phonetic analysis. This is true in spite of the fact that cuneiform was a highly complex form of writing and, as analyzed for the first time by Grotefend, was found to consist of a mixture of alphabetic writing, syllable recording, and pictographs. But here, as was the case with the Egyptian hieroglyphics, the multiplicity of pictorial representations was resolved into a much smaller number (approximately 250) of elements which thus could become the basic units for recording thoughts and documents. It might be noted in passing that Chinese has not carried through this stage of development even to the present day. Once the number of basic units had been reduced to a small number as in cuneiform or in the Egyptian hieroglyphics, the path to further development was open. It led—even though changes in further development have proceeded slowly—through the invention of printing (Gutenberg) to the individual letter as the basic unit for representing sounds, both vowels and consonants.

It is particularly important to note that these various processes for recording human thought by use of symbols have a common feature. The various forms of recording become effective as far as their subject contents and intellectual substance are concerned, only when they are transmitted through the human sense organs (eye, ear, sense of feeling with braille) and the human nervous system to the brain and thus regain their character as intellectual property of the human being. This is true regardless of whether the symbols are recorded on papyrus, engraved onto clay tablets which may be subsequently dried or baked, chiseled or engraved into stone, carved in metal and cast, drawn or printed on paper, or recorded by the various mechanisms of modern technology, either by the graphic arts or by methods of acoustical recording. Once the subject contents of documents have been “received” in the human brain, they become permanent possessions of the individual human being only if he is able to store such subject contents in his memory and, furthermore, only if he is able to recall such subject matter from memory, as needed, for comparison with other observations or acquired knowledge and for generating new correlations. This requirement is obviously closely linked to the magnitude and also to the finite capacity of the human brain, that is to say, to the capabilities and limitations of human memory as a storage instrument. We shall return to this consideration when we undertake the discussion of new possibilities for providing documents from storage and for evaluating their content.

Documentation—Its Scope and Definition

The attempts, initiated by Otlet and Lafontaine⁶, to accomplish for the first time a completely systematic coordination of the subject contents

of all written documents in bibliographic form provided the starting point for that type of activity which we today call "documentation." The term documentation was linked for the first time with that of an organization whose purpose it is to encourage and stimulate this activity with the establishment of NIDER (Nederlands Instituut voor Documentatie en Registratuur) in 1921⁷. Donker-Duyvis and Alingh Prins have thus contributed decisively to formulating the direction for development of documentation in future decades. The term documentation has been repeatedly defined in recent years. E. Pietsch and G. Mulert in 1954 surveyed⁸ opinions of leaders in librarianship and documentation as to the scope and definition of documentation. An analysis of the varying definitions was carried out in order to arrive at a generally accepted definition of the term. At the request of the FID, the Deutsche Gesellschaft für Dokumentation in 1954 drafted the following definition of the term: "To conduct documentation means, in a systematic fashion, to bring together documents, to analyze them and to render them useful. This activity is documentation"⁹. As is perhaps obvious, this definition takes into account various principal areas of activity by documentalists. As a consequence, this definition takes cognizance of the four fundamental operations which M. Hyslop cited as criteria for documentation¹⁰.

The Two Types of Documentation: Documentation of Written Records and Documentation of Experimental Facts and Observations

Two related yet distinct sources continually provide us with new observations, experimental facts and new knowledge in general.

The first such source involves direct sensory perceptions as provided, for example, by scientific experiments^{11,12}. Further examples of basic observational material are provided by the clinical reports, hospital records and the like of the medical profession^{13, 14, 15}.

The second general source consists of written records which are the result of human effort to express and to record in language the interpretation of observations, experiments and such, and also their theoretical analysis and synthetic expression, for example, in the form of scientific theory.

From a general point of view, it may be said that the accumulation of documentary material in the broad realm of intellectual activity that today finds its expression in science, technology, industry and related professional activity, compels us to develop methods which will guarantee the efficiency of intellectual activity at a level that will correspond to the highly developed status of technology in our time and to the further scientific and technological advances which can be confidently awaited provided the efficiency of intellectual activity can be insured as a basis for

such developments. The following areas require particularly careful attention.¹⁶

1. The working methods of individual investigators or intellectually creative persons who are confronted with the task of evaluating broad areas of diverse subject content in order to arrive at more penetrating understanding and broader correlations on the basis of available factual information.

2. The nature of intellectual activity, especially in the field of science, when a large number of parameters must be taken into account in carrying out an assignment. If each parameter involves a large number of individual facts which must be evaluated in order to arrive at higher order correlations then the case often arises that our human memory, even when supplemented by conventional documentation methods, is no longer able to provide the required ability to control and to correlate the actual material. A paradoxical situation is certain to result when technology develops and perfects new methods of investigation in various fields and thus permits us to approach and to study one and the same problem or experimental object from various points of view, and when a corresponding further development is not achieved at the same time in methods for evaluating and for correlating the facts provided by new investigative methods.

3. The continuing expanding volume of graphic records whose surveillance and evaluation becomes continually more difficult. In numerous cases, new facts and new knowledge recorded in the literature sink at the moment of their publication into the realm of the subconscious for long time periods. That is to say, newly acquired facts and knowledge are not incorporated into the general scheme of scientific understanding and thus made an effective factor in further development. All too frequently, a lengthy time interval intervenes before newly acquired facts and knowledge, perhaps through the medium of abstract journals or monographs, come to attention and are thus channeled into the main current of intellectual development. The practical consequence is that one and the same investigative study is repeated at different places without taking account of the fact that the problem being studied has already been investigated elsewhere. Loss of time and economic waste are unavoidable consequences.

The Accessibility and Retrieval of Recorded Knowledge as a Problem of Publication

As previously pointed out, when observational facts and related knowledge are rendered objective and recorded, that is to say, when documents are created, an intellectual sphere of human activity is created independent

of individual persons. At the same time, the question arises as to how a human investigator is to acquire those documents which have been made "available" when he needs them at a particular time that will be determined by his intellectual activity arriving at a certain point.

The first steps toward solving this problem were taken very early in the history of science. In ancient Greece, Aristotle¹⁷ pioneered in the scientific division of labor by systematically collecting and correlating not only his own scientific work but that of his outstanding pupils. We may thus regard Aristotle as the initiator of encyclopedic correlation as a tool for the advancement of science.

This important principle has continued to find recognition through the ages. Thus, in the first centuries A.D., Plinius¹⁸ in his "*Historia naturalis*" and especially Martianus Capella¹⁹, Boethius²⁰, and Isidor of Sevilla²¹ preserved the knowledge of the ancients in their writings and rendered it available to posterity. Special mention should be made of the early humanism²² of the thirteenth century, whose great encyclopedists belonged to the religious orders of Franciscans and Dominicans. Here we might mention the names of Bartholomaeus Anglicus²³, Thomas de Chantimpré²⁴, Vincent de Beauvais²⁵, and Konrad von Megenberg²⁶. In the eighteenth century period of intense intellectual activity, d'Alembert and Diderot²⁷ and contemporaries of similar intellectual tastes earned the title of "Encyclopedists." In our own time practically every large nation publishes one or more encyclopedias of knowledge; but these, in contrast to the encyclopedias of earlier times, are usually the work of numerous collaborators. In fact, they are often the work of large anonymous communities which, like ants, assemble bits of knowledge which form the building blocks for mammoth encyclopedic works. Today our knowledge is so varied in character and vast in bulk that no one person can comprehend it in the exhaustive fashion which was possible during the Middle Ages. After Leibniz²⁸ had mastered in a universal fashion the knowledge of his time, Alexander von Humboldt²⁹ was the last individual able to correlate the total knowledge of his time, the middle nineteenth century, into one comprehensive four-volume work: "*Kosmos*."

This development of an overwhelming volume of knowledge is a direct outcome of attention being directed to the individual phenomena of nature. This was urged by Roger Bacon in the thirteenth century; even though he was not understood by his own time, he nevertheless established the postulate "*sine experientia nihil sufficienter sciri potest*." Active application of this principal began at the time of Galileo (1564-1642)³⁰, whose postulate, to measure quantities capable of being measured and to render measurable any quantity incapable of being measured, has proved the cornerstone for the phenomenal development of all branches of natural science and of technology based thereon.

Kepler, Descartes, Newton, and an endless number of successors established such an abundance of closely related facts that it became more difficult from century to century, even from decade to decade, for any one individual to comprehend them all. This rapid expansion in factual information made it impossible for encyclopedias to encompass wide realms of knowledge. The creation of special compendia for individual sciences became necessary. This was done for medicine³¹ and chemistry, whose first comprehensive compendium, the "Alchemia" of Libavius³², appeared at the end of the sixteenth century shortly after the work of Paracelsus (1493-1541). This trend, once established, continued without interruption.

In the seventeenth and eighteenth centuries textbooks and compendia became more and more voluminous. Here again we might cite chemistry as an example³³. In this field, the comprehensive evaluation of the state of knowledge became particularly critical when all recorded facts had to be reorientated as a result of the overthrow of the phlogiston theory, especially by Lavoisier (1743-1794). An attempt to create a new textbook and compendium was undertaken by various writers; but in the course of the next few decades only one person was able to achieve impressive success. He was Leopold Gmelin in Heidelberg, who in 1817 published the first edition of his textbook of chemistry ("Lehrbuch der Chemie"). Before his death in 1857, Gmelin was able to initiate the fifth edition³⁴ of his textbook. His work has been continued down to the present day and has remained associated with his name^{35, 36, 37}, although organic chemistry has had, for many years, its own independent compendium, the "Handbuch der organischen Chemie," established by Beilstein³⁸.

In the nineteenth century there was a rapid advancement in the various fields of science. Consequently, the compendium, which by its very mode of publication must inevitably lag rather far behind the appearance of individual papers reporting new results, had to be supplemented in order to insure current awareness of the contents of the ever-increasing number of scientific periodicals. This need furnished the impetus for the establishment of various abstract periodicals. In 1830 the natural philosopher, Gustav Theodor Fechner, founded his *Pharmazeutisches Zentralblatt*, which from 1856 has been known to the professional world as *Chemisches Zentralblatt*³⁹. The importance of abstract periodicals to the advancement of science is clearly demonstrated by the fact that this one abstract periodical has in the meantime been followed by numerous others in the field of chemistry: thus in 1878 *British Chemical Abstracts* (at first within the *Journal of the Chemical Society*), from 1927 to 1954 separately as *British Chemical Abstracts* and in 1907 the *Chemical Abstracts* of the American Chemical Society⁴⁰. Other fields of science and technology have created their own abstracting services. To cite only one example: The Union Institute of Scientific and Technical Information in Moscow is now publishing sixteen

abstract journals to cover all fields of science and technology of interest to that Institute⁴⁶. Bradford, in his book, "Documentation," published in 1950 clearly shows how difficult the situation has become in spite of all these highly developed aids⁴¹. In Bradford's opinion the situation can only be characterized as a documentary chaos because "less than half the useful papers are noticed in the current abstracting and indexing periodicals."

Our brief survey of the development of scientific publication forces us to the conclusion, whose implications will be discussed later in detail, that the increasing volume of scientific publication has been the cause of increasing difficulties encountered in making available the factual information so essential to the advancement of science. These difficulties have been eased from time to time by the creation of compendia and abstract periodicals. Nevertheless, we are still confronted by a critical situation. It is perhaps worth noting that the crisis, as we face it today, occurs almost exactly five-hundred years after Johannes Gutenberg's invention of printing, that extraordinary accomplishment of which Carlyle said, "All that mankind has done, thought, gained or been is lying in magic preservation in the pages of books." But it has now become evident that what can be accomplished by printing is now approaching its limit, in fact, may have already reached it even today.

A clear indication of the urgency of this situation was provided by the World Congress for Documentation held in August, 1937, at the Maison de la Chimie in Paris⁴². Delegates were present from 45 countries. Some 30 governments and 40 international organizations were represented. The central questions were the following: What means do we have to master the steadily increasing flood of knowledge?. What approaches can we develop and utilize in order to accomplish methodical control of established facts? The formulation of these questions occurred some 20 years ago.

The extremely rapid evolution and development of the exact sciences and, as a consequence of technology and industry, have led to a crisis in publication⁴³ which finds its expression not only in the increase in the number of papers within a given periodical but also in the increase in the average length of the published papers (in spite of the efforts to achieve condensation by conscientious editors), as well as in the continuing increase in the number of periodicals within the various fields of specialization⁴⁴. As a consequence of these trends, the abstract journals in the form in which they have appeared during past decades are encountering increasing difficulties in striving to provide, within an acceptable time interval, abstracts of the various new papers. Very considerable expenditures and investments, both financial and in personnel, which have increased from year to year, have been required in order to maintain the preparation of abstracts. These increasing difficulties continue to stimulate the application

of improved and more efficient working methods. These statements apply to and are illustrated by the activities of the Chemical Abstracts services⁴⁶ in the United States as well as of Union Institute of Scientific and Technical Information in Moscow⁴⁶. The creation of a special research division at Chemical Abstracts in Columbus, Ohio, is an expression of the difficulties with which the flood of publication has confronted those organizations which produce abstracts⁴⁰.

Surmounting the unusual difficulties in conducting the examination of patent applications has been the subject of international concern and discussion. In 1955, on a certain day, approximately 200,000 patent applications awaited action in the United States Patent Office. Approximately $3\frac{1}{2}$ years passed from the filing of an application in the United States to the granting of a patent. The patent examiner had to devote approximately 60 per cent of his time to searching previously published material⁴⁷. Similar situations exist in other industrial countries.

In addition to further examples of such situations which were presented both in the first edition of this book and elsewhere, we might add some others which appear particularly important. In Figure 28-3 of the first edition of the book, estimates were given concerning the increase in published material pertinent to the Gmelin volume devoted to the element boron. The supplementary volume for boron appeared during the latter part of 1956. The exact data are now available. The main volume (Hauptband) which appeared in 1925 reported on the literature for boron during the preceding 150 years and cited 3,551 papers dealing with that element. The supplementary volume (Ergaenzungsband) which presented the literature until the end of 1949 and thus reported on 24 years of boron chemistry cited, on the other hand, 5,307 original papers. This, however, means, when one compares the 3,551 papers for 130 years and the 5,307 publications for 24 years, that the literature density for the time covered by the supplementary volume, that is to say, the rate of publication in the field of boron chemistry in the last 24 years, has increased by 800 per cent. To cite another example from the field of biology, we note that during the years 1861 and 1900, that is during a time interval of 40 years, 88 new periodicals appeared. In the time interval from 1951 to 1955, that is during 5 years, or $\frac{1}{8}$ th of 40 years, 290 new periodicals appeared—an increase of more than 26 times.

In addition to the periodical literature, which can be obtained by subscription, recent years have witnessed the appearance of a new more or less periodic form of publication, the report, whose documentary control presents additional difficulties. The librarian and documentalist are now in continual uncertainty as to whether there may be some organization or office—completely unknown—which is bringing out reports which should

be acquired and processed. Other difficulties include frequent change in character of such documents, unpredictable transfer from classified to declassified status, the supplementary publication of their content, in whole or in part, in one or another of the scientific or technical journals, and finally the necessity of maintaining continuing surveillance of such variations in the pattern of publication. Almost unnoticed by a majority of scientific and technical people, a new field of documentation has developed which requires its own methods of surveillance, acquisition, custodianship, and retrieval.

This state of affairs in the broad field of publication—viewed as a whole¹¹—has made it increasingly difficult to carry through an exhaustive and thorough search of the state of knowledge relative to a given problem or situation. It is becoming, as a consequence, more difficult to meet the requirement that a new assignment will be undertaken on the basis of awareness of all previously acquired pertinent facts and knowledge. Cases become more frequent, as a result of which assignments in research and development are undertaken and then the observation is made, either in the course of carrying through the assignment or after its conclusion, that the work had already been accomplished previously in the same or in a very similar fashion and that, consequently, both the time of skilled personnel as well as associated expenditures had been completely, or to a large degree, wasted. On the other hand, achieving access to previously recorded knowledge buried in our continually expanding libraries requires an extraordinary application of bibliographic aids quite aside from the fact that the accomplishment of literature searches is becoming more and more difficult since their performance requires the person carrying them out to have not only a knowledge of the subject matter but also searching experience of such a nature that it can be achieved only by special schooling or extensive experience. As a consequence, the proposal is repeatedly made not to bother with the Sisyphus-like task of conducting a literature search before undertaking to carry out a new assignment in research or development^{48,49}.

The situation, as outlined above, however, must unavoidably lead from year to year to a rapidly sinking level of efficiency in the field of intellectual accomplishment—and this quite independently of the economic consequences. A further important difficulty is the fact that human memory—considered as a storage organ—is confronted with increasingly difficult tasks which bring it to the limit of its capability. In fact, human memory in numerous cases is already no longer able to cope with information as it becomes available both because of its amount and also often because of the rate of accumulation of factual material. Thus human memory exhibits a natural limitation due to the physiological structure of

the human brain. This limitation compels us to seek a solution to the problem of the storage of facts and knowledge in the creation of an extra-human storage aggregate functionally similar to memory^{50, 51}.

The well-established fact of overloading the memory of intellectually creative persons has, indeed, stimulated investigations as to whether and to what degree it is possible to create storage aggregates of extra-human character⁵². These storage aggregates would have the function of providing, on demand, factual material as needed for any particular assignment or problem. In other words, the storage aggregates would provide the necessary material for carrying through creative thought without burdening human memory. The broad range of factual material to be found in the subject contents of documents as well as the wide range of knowledge continuously generated by scientific research would find in such extra-human storage aggregates a reliable and objective preservation.

A number of requirements must be made of these storage aggregates. They must operate at a rate which approximates that of human memory. They must permit an extremely wide possibility of combining units of information or knowledge. On the other hand, they must be free of the various limitations of human memory. In particular, they must be able to accept rapidly large volumes of factual data or recorded knowledge without exhibiting fatigue and without the tendency of human memory to forget.

The various forms of external, extra-human, storage aggregates that either have been already developed or are technically feasible at the present time for storing the factual data and knowledge in documents may be subdivided as follows:^{11, 48, 53, 54, 55, 56}

1. mechanically operating devices
2. electro-mechanically operating devices
3. photoelectrically operating devices
4. electronically operating devices.

All of these devices have one thing in common—they record signals to which meaning is assigned in the sense of information theory. In other words, concepts and semantic units are recorded in a form which is unambiguous, free of contradiction, logically precise, particularly with regard to related concepts, that is to say, in a "spectrum pure" form as so-called documentation elements. Such recording is carried out in a sharply defined relation on topologically sharply localized regions of the storage media, that is to say, on elements characterized by surfaces. In other words, the subject content to be stored in extra-human memories is transferred with the aid of symbols from previous forms of recording onto some surface or other as exemplified by cardboard, metal foil, film, or the surface of a magnetic drum. This act of recording is accomplished so that, of itself, it

does not result in distortion of the subject contents which are preserved and stored until, at a given moment by an inverse process of recall, they act to bring to human attention the factual material originally involved in the documents as processed prior to storage.

The proposal to apply automatic processes for documentation and selection still encounters widespread misapprehension. The principal cause of concern appears to be the feeling that intellectual creativity is somehow threatened by the new methods. Their rejection is, however, at least to a considerable degree, the result of a stand pat attitude which is little short of astonishing, when it is considered that intellectual activity has achieved revolutionary progress in many fields but that at the same time, the methods which serve and expedite intellectual accomplishment itself remain traditional in character even today. Quite apart from such conservatism, another difficulty arises from the fact that it is very difficult—at least in the early stages of development—to prove that modern methods of documentation will yield economic advantage. But perhaps the basic reason for the rather critical situation of modern documentation methods lies deeper and concerns the character of automatic documentation itself.

In this connection, it must be recognized that various machines and types of equipment which have been applied to documentation were developed and designed originally for quite different operations, especially in the field of accounting, statistics, mathematics, and the like. It is to be expected, consequently, that a more or less extensive development cannot be avoided if such equipment is to be applied to documentation, i.e., if it is to be applied for recording and selecting quite complicated forms of subject matter and knowledge. Machines must be designed to meet the particular requirements of documentation.

What is true with regard to machines and equipment is also true, even more emphatically, with regard to the forms of expression and related symbolism with which the results of the analysis of the subject content of documents will be recorded in the storage aggregates.

It must be kept in mind that both of the above mentioned aspects of the mechanical documentation development constitute a radically new type of problem which is subject to its own laws and principles. If we attempt to work with the previous methods of indexing, classifying, and coding, we should not expect that such methods will be adequate for carrying through machine operations. If we attempt to apply the old methods, we should not be surprised if the results obtained do not provide the advantages that had been anticipated. These considerations provide the starting point for a new development along the following lines. First, account must be taken of the requirements which are imposed on documentation methods by developments in research and experimental techniques and the resulting expansion in recorded facts and new knowledge. If we do this, we come to

the conclusion that new methods for organizing and analyzing the subject content of documents must be established and that such methods must be aligned with the new requirements on the one hand, and the possibilities offered to us by automatic equipment, on the other hand. The importance of this point can scarcely be overemphasized. In this connection, it should be clearly stated that the investigation and the development of such new methods does not mean that we underestimate or even discard previous documentation methods which have been found in practice to provide advantageous results. However, recognition of this fact should not blind us to the nature of the problem of developing new methods. It is gratifying to observe in this connection that it is becoming more generally recognized that an essential prerequisite to achieving a satisfactory solution of the present problems in documentation is the development of appropriate principles for organizing and analyzing factual information in the various areas of professional and intellectual activity together with the application of such methods for storage and in connection with automatic devices. Recent trends in documentation research point quite clearly in this direction and, in fact, the trend has become so well developed that it is no exaggeration to state that a new field of research has been opened up. It is particularly fortunate in this connection that the question of the relationship of man to his machines has attracted interest in a more fundamental way and on a broader basis than was the case during the nineteenth century. Investigations along these lines inevitably involve the study of the memory capacity of the brain and are of unusual interest for documentation and should attract the full attention of documentalists. This statement remains true in spite of the fact that any suggestion of devaluing creative human intellectual activity is to be opposed or at least to be viewed with the greatest concern and reserve⁵⁷.

It must be emphasized that we cannot be satisfied in the field of documentation with partial solutions, but rather that we must grant to this field the status of a new area of research and in particular, as described above, we must direct attention to two aspects of further development. One pertains to the area of machines and the second to the development of new methods and, in particular, the formulation of analytical procedures for recorded facts and knowledge, not only with respect to the concepts and symbols used for their expression but also from the point of view of mathematical logic.

The Gmelin Institute and Automation of Documentation Procedures

Attention will now be focused on the effort and work that has been performed at the Gmelin Institute for Inorganic Chemistry and Related Fields for the purpose of increasing the over-all efficiency of the Institute

staff by carefully studying and introducing, step by step, automation of various operations. The first edition of this book presented a progress report of the first attempts along these lines. During recent years the further development of new methods has passed through various phases. A policy of restraint was followed as it was recognized that it would be advisable—and this is perhaps a valid general rule for the processing of large volumes of documentary material—to avoid premature or hasty transfer of a large volume of recorded information, for example, an extensive information archive, into the new mechanical form even though it had become evident that a workable set of methods had been developed. If one considers the experiences of many documentation organizations in recent years, it becomes clear that the effectiveness of the new methods has been repeatedly proved in different fields of specialization. At the same time, it has become evident that, from such favorable experience, it will be possible to derive more general and broader principles which can guide and control further future development. As far as Gmelin is concerned, it may be stated that the new methods have achieved a useful, well-established formulation so that it is now possible to transfer broad areas of Gmelin's general field into automatic form. This has been accomplished at present for the minerals. The transfer process is under way for alloys of nonferrous nature. Information files relating to the individual element chromium are also undergoing reprocessing from the classical card catalog to machine-sorted punch cards. Preliminary work is under way for an even larger area, namely, that of the alloys of iron including the steels.

In the first edition of this book, it was pointed out that the classical handbook as exemplified by Gmelin was entering an era of uncertainty. It would be impossible to disregard the frequently expressed opinion that the era of the handbook as a means for reporting and summarizing the status of a broad field has passed.

On the other hand, it is scarcely necessary to emphasize that an Institute such as Gmelin would devote considerable time and attention to careful and methodical study of the question of the best methods of presenting in documentary form the information within its field. This study has involved experimental investigations of various processes. Before considering the results of these studies, it should be stated, to avoid any misunderstanding, that the 8th edition of the Gmelin handbook, which has been in preparation since 1921–22, will be continued and that a 10–12 year plan has been worked out to publish the completed 8th edition in book form. The closing date for the literature for this 8th edition was January 1, 1950. In carrying through this plan, it is not the form of the handbook which has undergone change, but rather the working methods of the Institute staff. These changes

have involved, in particular, the preparatory processing of the Institute's documentation center. The goal of such changes has been to improve the efficiency of the staff as a whole, but most especially those staff members who are scientifically trained and who are responsible for writing the handbook text. In working toward this goal, there have been extensive changes in the preparatory processing, that is to say, in the aids that are provided to support the intellectual activity of the scientific staff. Their task remains, as before, to review the original literature in exhaustive and, at the same time, critical fashion, that is to say, to read each original publication relating to a given section or subsection of the handbook and to prepare the text for printing. In this connection it should be pointed out that a 9th edition of the handbook in traditional form will be written. The most modern methods of automatic documentation will be applied for processing the literature, in order that various operations necessary in preparing the 9th edition of the handbook may be performed with top efficiency. From the broader point of view of methods for correlating knowledge and presenting such correlations, it appears particularly important to point out that for this broad area of science, the state of knowledge will be presented in handbook form based on archives generated to provide the needed factual background. In writing the handbook in the future, the work methods must be so developed that the publication of the handbook can be accomplished with a lessened lapse of time—this requirement constitutes reason enough why the Gmelin Institute should be concerned not only with applying new methods for increasing the efficiency of the staff but also with the future development of such methods.

The Gmelin Institute—Its Mission and Its Status in 1958

The Gmelin handbook was founded by Leopold Gmelin in 1817³⁴. Since 1921–22, a staff of full-time scientific personnel, which has been frequently expanded, has been working^{35, 36, 37} on the 8th edition which will present a complete review of the field of inorganic chemistry and bordering fields from the time of the beginning of modern chemistry with the abandonment of the phlogiston theory down to January 1, 1950. This summary review embraces the total world literature and is organized on the basis of the modern principles of physical chemistry. More specifically, the areas embraced by the handbook are: inorganic and physical chemistry, furthermore analytical chemistry, electrochemistry, experimental physics (insofar as individual substances are concerned, the nucleus, the atom, the ion, the molecules, electrical properties, magnetic properties, mechanical properties, mechanical-technical properties, optical properties, radioactivity, thermal properties), preparative inorganic chemistry, geochemistry and bio-geo-

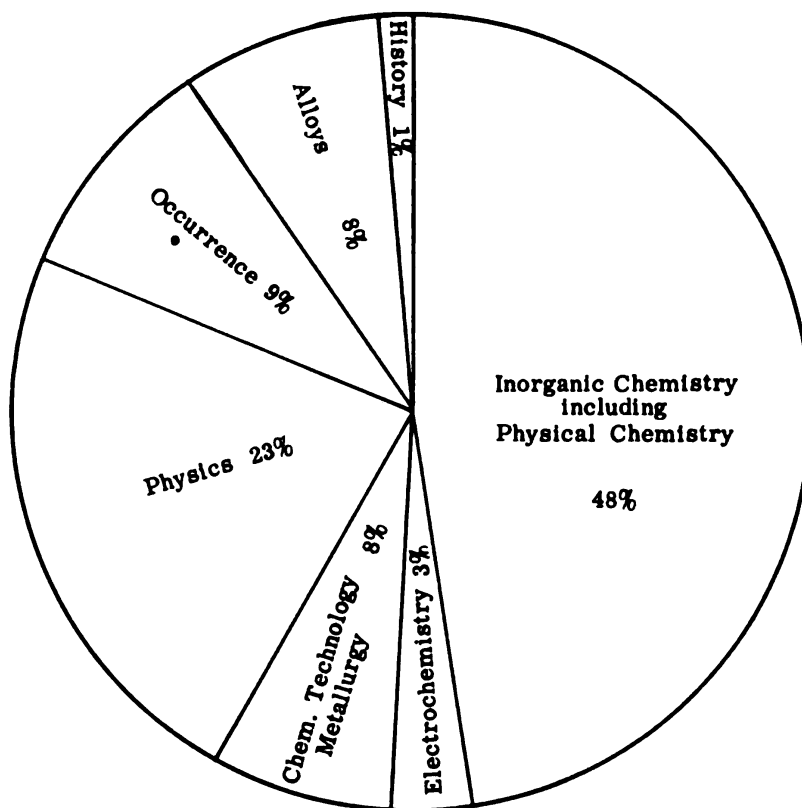


Figure 28-1. Percentage of papers in areas of specialization covered by Gmelin.

chemistry, geology, history of chemistry, heterogeneous equilibrium, isotopes, colloid chemistry, corrosion, crystallography, ores and ore deposits, metals (noble metals, light metals, other metals), metallography, metallo-organic compounds, metallurgy, mineralogy, surface protection, passivity, physiological behavior (industrial hazards and counteraction), technology, chemical industry and related statistics. See also, the main divisions of the Gmelin subject matter system, pages 594–595.

Preparation of the handbook text is based on extensive and complete archives which include the literature to January 1, 1950. Preparation of the text involves a critical study of the original literature. Figure 28-1 shows the extent to which various areas of specialization, such as electrochemistry, constitute a part of the total area embraced by Gmelin. These percentages are computed on the basis of the number of papers in the various fields.

The material, that is to say, the factual information relating to chemical elements and their compounds, insofar as it does not fall within the field of the Beilstein handbook of organic chemistry, is arranged in accord with the so-called Gmelin principle of the latest position which is summarized and illustrated in Figure 28-2. In accordance with this principle, the elements

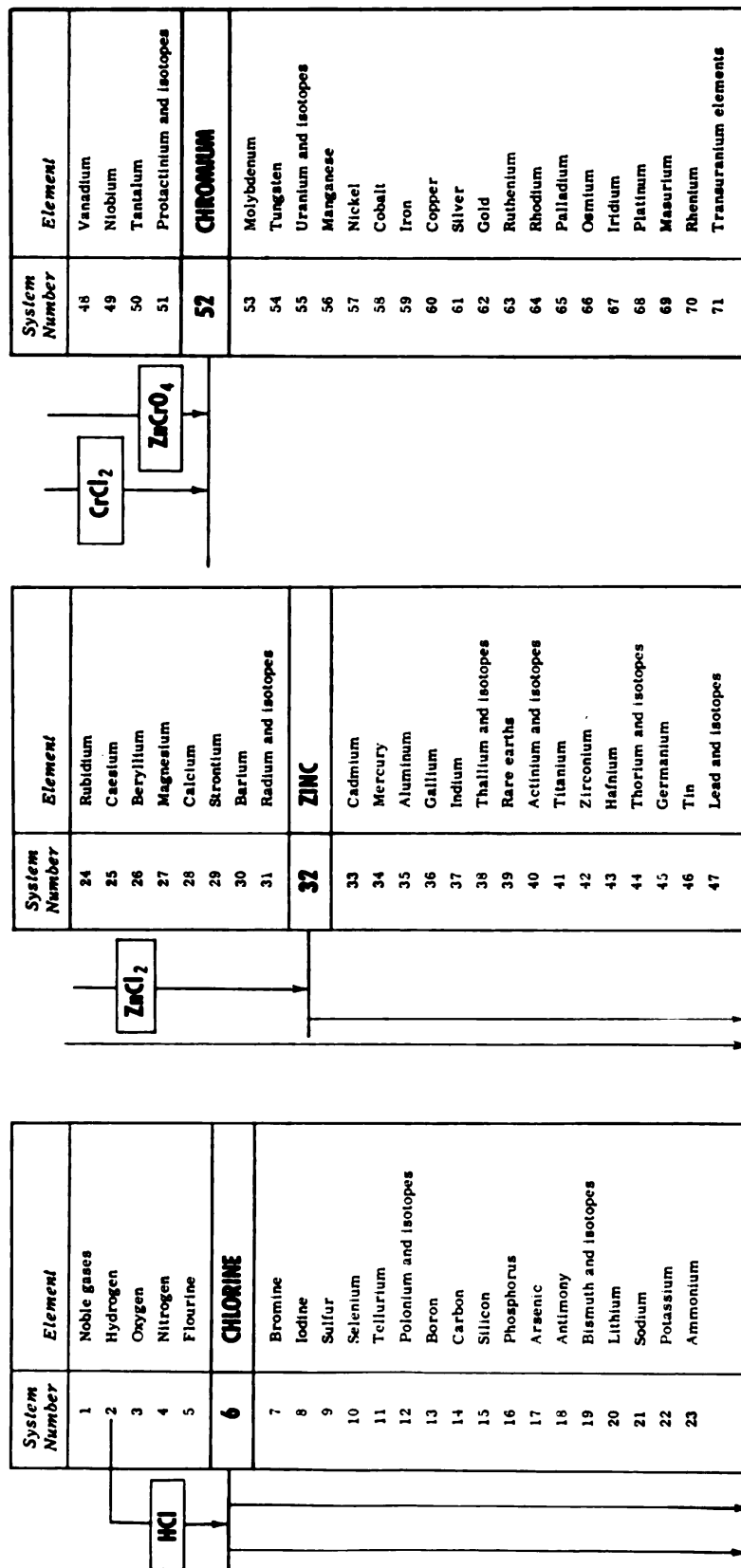


Figure 28-2. Ordering of elements and compounds by the Gmelin system as employed in the 8th edition of the Gmelin Handbook. (Compare also the expanded form of this system in Figure 28-5 as adapted to automatic documentation by Gmelin).

and their compounds are arranged in a fashion different from the periodic system. The Gmelin system is characterized essentially by the fact that elements which form anions precede those which form cations. The system operates in such a way that the characteristic compounds of a given element are described under that element's system number. Thus the volume of the handbook for an element with the system number n presents all the compounds and combinations of this element with all other elements whose system number is less than n , that is to say, all elements with system numbers 1 up to $n-1$. For example, the system number 59-iron, presents all known combinations of iron with the element 1 (noble gas) up to 58 (cobalt). For any compound or combination, the component element with the highest system number determines the assignment to a particular volume. For example, Fe_2O_3 will be found under the system number 59-iron, and not under the system number 3-oxygen. Furthermore, Pt_2Fe will not be found under the system number 59-iron, but under the system number 60-platinum. If a given compound consists of three or more elements, then this compound will be found in the volume for the component element with the highest system number involved and within this volume the next lowest system number is decisive. For example, rubidium bromochloride is found in the volume for rubidium (system number 24 under rubidium and bromine) while rubidium bromoiodide will be found under rubidium and iodine.

During the interval 1924 to 1956, the Gmelin Institute has published 41,973 pages of printed text in 143 volumes or pamphlets (Lieferungen) and also has published 8,825 pages of the so-called Gmelin Patent Collection. Time of publication of various items and also the rate of publication are given in Figure 28-3. In 1950, a plan was worked out for completing the work on the 8th edition. This plan could, however, come into full effect only after surmounting considerable economic difficulties, that is to say, since 1955-56. According to this plan, the 8th edition will be completed within 10 to 12 years. The literature coverage of those volumes in the 8th edition which appeared before 1950 will be extended by supplementary volumes which will summarize the literature for the intervening period up to the literature closing date of January 1, 1950. Once this over-all plan has been accomplished, the 8th edition of Gmelin will provide a complete summarization of approximately 250 years of scientific and technical effort in the field of inorganic chemistry and related fields. This presentation will be based on exhaustive and critical evaluation of the total literature of this broad field from the point of view of present day prevailing theories.

At the present time (July, 1958) the staff of Gmelin consists of the following personnel: 55 scientists (included here is the Institute management, chiefs of divisions, and editorial personnel), 70 scientific-technical staff members, clerical and technical aids (including administrative aids). Thus, at the above mentioned date, the Gmelin staff numbered 125 persons.

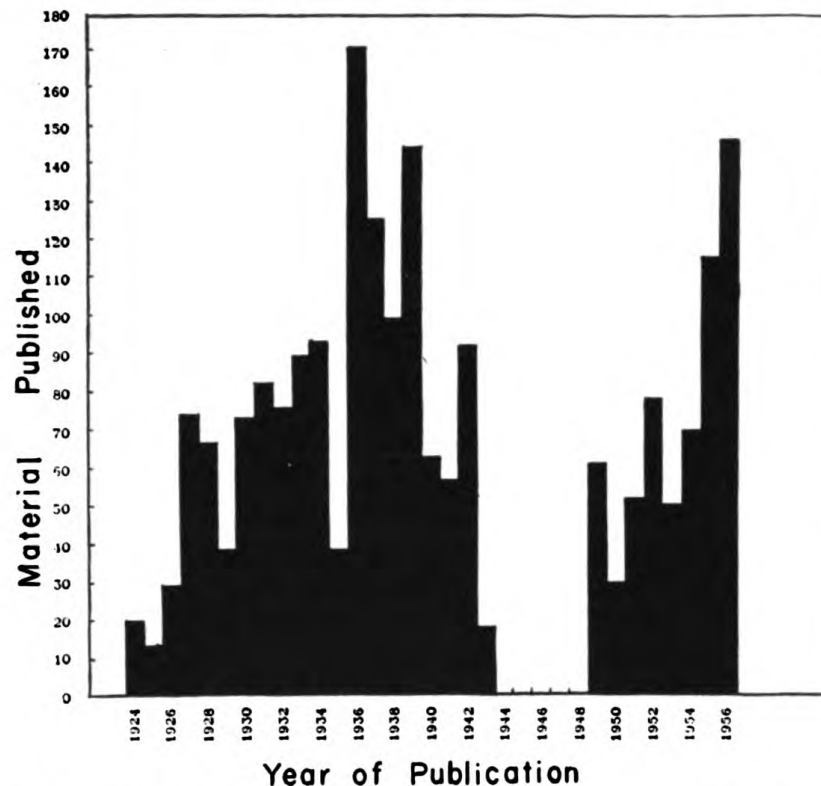


Figure 28-3. Amount of material published in successive years.

One of the principal tasks involved in planning and guiding the work of the Institute is to insure that the scientists on the Institute staff can devote full time to their intellectual tasks. The scientists are not only the intellectual backbone of the Institute but, from an economic point of view, they constitute the principal source of expense in operating the Institute. The total salaries for scientific and scientific-technical, and technical personnel, including management and administration, amounted to approximately 76 per cent of the total Institute budget. It is consequently particularly important to create working conditions in which the Institute scientists can devote themselves completely to the study of the original literature, to its comparative evaluation, and to the creative task of writing the text for the handbook without being deflected from these tasks by any auxiliary or preliminary processing of the input information. Particularly important preparatory operations include the following: the collection and processing of the literature which is to be used by the scientists in conducting their work as outlined above, provision of books and of photocopies and microfilms, technical processing of the manuscript preparatory to setting type, as well as proofreading of galleys, page proof, etc., in publishing the various volumes of the handbook. These various tasks are accomplished by special working groups that function as supporting sections within the Institute's

over-all organization. Particularly important for the purpose of insuring and improving the efficiency of the scientific staff members are the following:

1. the Institute's library
2. the section for documentation (subject archives)
3. the section of document reproduction
4. the section for technical processing of manuscripts, for proofreading and similar work.

During the years immediately following World War II, which resulted in the total destruction of the previous Gmelin library, the library staff was compelled to devote much of its time to arranging for inter-library loans. Since 1949, the Institute has been a member of the German Interlibrary Loan System and in this way is linked with 320 libraries both inside and outside Germany. Loan operations, which on an average involve about 3,000 book loans per month, have been greatly facilitated and accelerated by the creation and maintenance of an extensive central catalog by the Institute's library staff. It has proved possible gradually to develop the Institute's own library whose present holdings (July, 1958) might be summarized as follows: around 20,000 volumes (monographs and bound periodicals), 199 continuing subscriptions to German periodicals, and 376 subscriptions to non-German periodicals, continuing acquisition of the brochures of 1,500 industrial firms (particularly American), approximately 25,000 reprints (on a subscription basis), 152,000 pages of microfilms of scientific papers including around 7,000 patents.

The task of the documentation section of the Gmelin Institute may be summarized as follows: to establish and to maintain an archive collection as complete as possible of the literature necessary for the Gmelin handbook and, in particular, without limitation to the literature closing date of the current 8th edition. In other words, processing of the literature by the documentation division continues on a current basis and, in this way, the documentary basis for a later 9th edition of the handbook is being provided.

The documentation division of the Institute continues to generate archives in the "classical form" even though, as explained subsequently, it is also working on methods for applying automatic documentation.

The literature is processed according to the Gmelin key system (explained in detail later) which is set up to correspond to the subdivision of the handbook. This processing results in preparing archive cards arranged according to author, bibliographic citation, literature source, subject contents. On an average, each publication that is processed requires three cards. When a scientific staff member undertakes the preparation of the text for a section of the handbook, he is supplied with the corresponding archive cards as the basis for his work which involves as an essential step the study of the origi-

nal literature. The latter is supplied to him on demand by the Institute library.

The various archive cards are prepared in a variety of ways:

- (a) by direct evaluation of the more important periodicals of the world (this constitutes a major part of the work of the documentation division),
- (b) by evaluating appropriate abstract journals,
- (c) by evaluating various special publications in certain scientific and technical fields,
- (d) by evaluating reprints which are sent to the Institute on a regular basis from various countries and which constitute a very valuable supplement especially for the less common publications,
- (e) by taking into account the subject contents of informative publications of the various industrial concerns,
- (f) by review of various reports and other unpublished literature, sometimes referred to as "underground literature" whose acquisition requires particular attention and effort.

Approximately 1,538,200 cards were in the archives of the documentation division in July, 1958. The so-called Gmelin key system, which in recent years has been further developed and systematized for the purpose of using it as a basis for automatic documentation, has been developed by the Gmelin Institute staff in a systematic fashion for about the last 35 years. The usefulness and practicality of this system are being continually tested by the day-to-day work of the Institute. As already noted, this key system corresponds in all details with the arrangement and systematization of the subject contents of the Gmelin handbook itself. The Gmelin key system as published⁵⁹ in 1957 had repeatedly proved itself in various applications as free from inner contradictions with respect to the scientific field which the system is called upon to embrace in exhaustive systematic fashion. It should be emphasized that this key system could be expressed in the form of a 3- or 4-letter code if this were required to achieve efficiency of storage with certain forms of automatic equipment

In addition to the systematic arrangement of the handbook itself, the factual material is analyzed on the basis of a more extended and alternate arrangement of the elements. This permits, for example, certain individual elements within groups as, for example, the elements of the rare earth series, to be individually considered. At the same time, various element groups, for example, the alkali metals, which up to the present time have not been essential to the handbook system itself, are accorded due consideration. During the evaluation and processing of the original literature, the various recorded facts in the original publications concerning a substance are arranged under two principal characteristic features. (1) The substance itself, is characterized as to chemical composition and physical form and, (2)

as to its chemical and physical properties, that is to say, with regard to its behavior insofar as it is known. These two features may be summarized as follows:

A. Substance arranged according to principal element, principal substance, physical condition of substance.

B. Factual subject matter arranged according to principal division, main group, sub-group, specific heading.

It is possible within the limits of this chapter to give no more than a general review of the principles underlying the system for organizing factual subject matter in the Gmelin compendium. An enumeration of all of the main divisions provides an indication of both the extent and also the limitations of the range of subject matter covered in the handbook.

The main divisions of the Gmelin subject matter system have been worked out as follows and arranged in the following order:

- 01 general literature
- 02 historical
- 03 physiological behavior
- 04 applications
- 05 economic aspects
- 06 mining methods
- 07 processing
- 10 analysis (in general)
- 11 analysis, qualitative
- 12 analysis, qualitative (relating to special materials)
- 13 analysis, quantitative
- 14 analysis, quantitative (for special materials)
- 15 analysis, quantitative separations
- 20 occurrence (general)
- 21 extraterrestrial occurrence
- 22 geochemistry
- 23 ores and ore deposits
- 24 minerology
- 29 systems
- 30 formations
- 31 preparations (in the laboratory, technical preparation)
- 32 preparation and production (involving water)
- 33 further processing
- 34 defects and errors
- 35 surface treatment and corrosion protection
- 40 physical properties in general
- 41 atomic nuclei (properties of the atomic nuclei)
- 42 atoms (properties of atoms and of element-ions)
- 43 molecules (properties of the molecules)
- 44 crystallographic properties
- 45 mechanical properties
- 46 thermal properties
- 47 optical properties
- 48 magnetic properties

- 49 electrical properties
- 50 electrochemical properties
- 60 chemical behavior (and corrosion)

To indicate how the system for organizing subject content is worked out on a detailed basis, an arbitrary example has been chosen; main division 13: Analysis, will be presented.

- 13 Analysis, quantitative
- 13.00.01 gravimetric (methods)
- 13.00.01.A gravimetric analysis (methods)
- .B thermogravimetric (methods)
- electrogravimetric (methods)
- 13.00.10 volumetric (methods)
- 13.00.11 indicator methods (general)
- 13.00.12 neutralization methods
- 13.00.12.A acidimetric (titration)
- .B alkalimetric (titration)
- 13.00.13 oxidation reduction methods
- 13.00.13.A manganometric (titration)
- .B bromatometric (titration)
- .C iodometric (titration)
- .J bromometric (titration)
- .K cerimetric (titration)
- .L vanadometric (titration)
- 13.00.14 precipitation methods
- 13.00.14.A argentometric (titration)
- 13.00.15 complexometric methods (chelatometric methods)
- 13.00.19 (further indicator methods)
- 13.00.20 potentiometric (titration)
- 13.00.21 conductometric (titration)
- 13.00.22 amperometric (polarometric titration)
- 13.00.23 coulombometric (titration)
- 13.00.24 high frequency titration
- 13.00.25 optical (titration)
- 13.00.26 calorimetric (thermometric titration)
- 13.00.27 cryoscopic (titration)
- 13.00.28 viscosimetric (titration)
- 13.00.29 stalagmometric (titration)
- Supplement:
- 13.-.-.1 microanalysis
- .2 semimicroanalytic
- .3 semiquantitative
- .4 apparatus

This system, developed for the subject contents of papers, has been summarized as to its essential features. The system, as already noted, is the basis not only for the "classical" Gmelin card catalog which provides to the scientists the basic material for undertaking the working out of the various sections of the handbook. See Figure 28-4. (In other words the card catalog provides the scientist with references. In every case, that is to say,

		<u>1953</u>
	<u>Si</u>	<u>SiCl₄, Prepn., chemical method, reaction</u>
	Si	Silicon, crystalline, chemical behavior with salts of inorganic acids Chlorides, CuCl
	Cu	CuCl, solid, chemical behavior, with non-metals
3-w-C-821		

		<u>1953</u>
	Si	SiCl ₄ , Prepn., chemical methods, reaction
	<u>Si</u>	<u>Silicon, crystalline, chemical behavior</u> <u>with salts of inorganic acids</u> <u>Chlorides, CuCl</u>
	Cu	CuCl, solid, chemical behavior, with non-metals
3-w-C-821		

		<u>1953</u>
	Si	SiCl ₄ , Prepn., chemical methods, reaction
	Si	Silicon, crystalline, chemical behavior with salts of inorganic acids Chlorides, CuCl
	<u>Cu</u>	<u>CuCl, solid, chemical behavior, with non-metals</u>
3-w-C-821		
Reaction in solid states. III. The temperature of reaction between metallic silicon and cuprous chloride. Teruichiro Kubo and Wazo Komatsu (Tokyo Inst. Technol.). <i>J. Chem. Soc. Japan, Pure Chem. Sect.</i> 74, 709-12(1953).—The mixts. of Si and CuCl were heated in const. vol. and the pressure of SiCl₄ resulting from the reaction as well as the temp. of the mixt. were measured. From the results the lowest temp. of reaction was estd. to be 230°. T. Katsurai		
3-w-C-821	C.A.	<u>1954</u> 3120

Figure 28-4. Gmelin file cards.

for each archive card, the scientist must then turn to the original literature.) The system as outlined above, also provides the basis for the Gmelin system for automatic documentation which uses machine-sorted punched cards of the IBM type.

Research on Automatic Documentation in the Gmelin Subject System

As already noted, the automatic documentation processes as developed by Gmelin are based on the subject matter system previously outlined. This

system has been developed for the field in which Gmelin is interested principally, namely, inorganic chemistry. In this field, it is particularly necessary to take into account the chemical formulas and specifically to provide their entry on the cards in a linear form. Various fundamental considerations underlying the Gmelin system of automatic documentation were presented in the first edition of this book. Since its publication, the Gmelin Institute in cooperation with Hollerith Division of the Max-Planck-Gesellschaft, Göttingen, has carried through extensive experiments⁶⁰ and, as a consequence, the methods and procedures as presented in the first edition of this book have undergone further refinement and extension.

The requirements which determine how formulas and factual information are recorded in machine-sorted punched cards may be summarized as follows:⁶¹

1. The complete characterization of the chemical substance concerning which a factual statement is made.
2. Recording of the factual information in a fashion which is not ambiguous as to characterizations and, as far as possible, relates to individual facts.
3. Precise citation of the literature reference pertaining to the card entries as summarized by 1 and 2 above.

With regard to characterizing individual chemical substances, that is to say: elements, compounds, alloys, polycomponent systems, and minerals, the following considerations are believed to be of decisive importance.

Two requirements are to be anticipated on the part of persons directing inquiries to the documentation system:

(a) It should be possible to direct an information request to each component within a compound or within a mineral. Thus, it should be possible to select from the totality of minerals, for instance, all those which contain germanium or it should be possible to select all chlorides or phosphates from the totality of known and encoded inorganic compounds. This requires that each chemical individual—as noted above—shall be regarded as consisting of its various components and each of these must be recorded by a particular hole combination in the punched cards.

(b) It is necessary to cite for the person requesting information the chemical individual in a readable form. The conversion of the punched-card language to readable form may be accomplished by passing the cards through a tabulating machine which then prepares the desired listing.

The requirement cited under (a) above may be carried out by the assignment of numbers to element symbols. In doing this, the Gmelin classification deliberately selects as its starting point, not the periodic system number of a given chemical element, but its Gmelin number as assigned and applied in the above discussed Gmelin principle of latest position. See Figures 28-2 and 28-5. Every compound or element combination in alloys, polycomponent systems, minerals, etc., contains some one chemical element whose

Gmelin Documentation Code Numbers for Elements and Groups of Elements Expanded Gmelin-System

a1		Noble gases	49	Ce	Cerium
01	He	Helium	50	Pr	Praseodymium
02	Ne	Neon	51	Nd	Neodymium
03	Ar	Argon	52	Pm	Promethium
04	Kr	Krypton	53	Sm	Samarium
05	Xe	Xenon	54	Eu	Europium
06	Rn	Radon	55	Gd	Gadolinium
07	H	Hydrogen	56	Tb	Terbium
a8		Non-Metals	57	Dy	Dysprosium
08	O	Oxygen	58	Ho	Holmium
09	N	Nitrogen	59	Er	Erbium
A0		Halogens	60	Tm	Thulium
10	F	Fluorine	61	Yb	Ytterbium
11	Cl	Chlorine	62	Lu	Lutetium
12	Br	Bromine	63	Ac	Actinium
13	J	Iodine	64	Ti	Titanium
14	At	Astatine	F5		Heavy metals
15	S	Sulphur	65	Zr	Zirconium
16	Se	Selenium	66	Hf	Hafnium
17	Te	Tellurium	67	Th	Thorium
18	Po	Polonium	68	Ge	Germanium
19	B	Boron	69	Sn	Tin
20	C	Carbon	70	Pb	Lead
21	Si	Silicon	71	V	Vanadium
22	P	Phosphorus	72	Nb	Niobium
23	As	Arsenic	73	Ta	Tantalum
B4		Metals	74	Pa	Protactinium
24	Sb	Antimony	75	Cr	Chromium
25	Bi	Bismuth	76	Mo	Molybdenum
B6		Alkalies	77	W	Tungsten
26	Li	Lithium	78	U	Uranium
27	Na	Sodium	79	Mn	Manganese
28	K	Potassium	H0		Metals of the iron group
29	NH ₄	Ammonium	80	Ni	Nickel
30	Rb	Rubidium	81	Co	Cobalt
31	Cs	Caesium	82	Fe	Iron
32	Fr	Francium	83	Cu	Copper
C3		Alkaline earths	H4		Noble metals
33	Be	Beryllium	84	Ag	Silver
34	Mg	Magnesium	85	Au	Gold
35	Ca	Calcium	H6		Platinum metals
36	Sr	Strontium	86	Ru	Ruthenium
37	Ba	Barium	87	Rh	Rhodium
38	Ra	Radium	88	Pd	Palladium
C9		Non-ferrous metals	89	Os	Osmium
39	Zn	Zinc	90	Ir	Iridium
40	Cd	Cadmium	91	Pt	Platinum
41	Hg	Mercury	92	Tc(Ma)	Technetium (Masurium)
D2		Light metals	93	Re	Rhenium
42	Al	Aluminium	I4		Transuranium
43	Ga	Gallium	94	Np	Neptunium
44	In	Indium	95	Pu	Plutonium
45	Tl	Thallium	96	Am	Americium
D6		Rare earths	97	Cm	Curium
46	Sc	Scandium	98	Bk	Berkelium
47	Y	Yttrium	99	Cf	Californium
48	La	Lanthanum	100	E	Einsteinium
			101	Fm	Fermium
			102	Mv	Mendelevium

This expanded form of the Gmelin System is designed for automatic documentation. The code numbers, as given here, (with each element assigned a number) do not correspond to the system numbers of the Gmelin Handbook, in which, for example, the noble gases are cited under a single number. The order of citation of the elements, as given above, is, however, the same as in the Handbook.

Figure 28-5. Expanded Gmelin system as adapted to automatic documentation.

number is, of course, identical with that which is decisive in determining the latest position in the Gmelin system. The element so characterized by the highest system number within a given element combination is chosen as the leading element in the Gmelin punched-card documentation.

With respect to the leading element, all of the other elements in such a combination are regarded as following elements. They are cited in order of decreasing system number and these system numbers in turn serve as the basis for the further characterization of chemical composition. No more than five such numbers are punched in the machine-sorted punched card. In this way each compound is entered in the punched card by punching the numbers for the leading element and, at most, five following elements. These numbers are punched in the so-called number field of the card. It should be pointed out that this form of characterizing a chemical substance in the number field consists solely of indicating the component elements in the order as determined by their Gmelin number. Each element in a given combination is mentioned or punched in the card only once. Furthermore, the stoichiometric composition of the chemical combination insofar as multiplicity of the number of atoms per molecule or similar stoichiometric proportions, e.g., of alloys, is not entered in the card. Such stoichiometric characterization of the element combination in the form of an individual formula is achieved, however, in the so-called alpha field of the card.

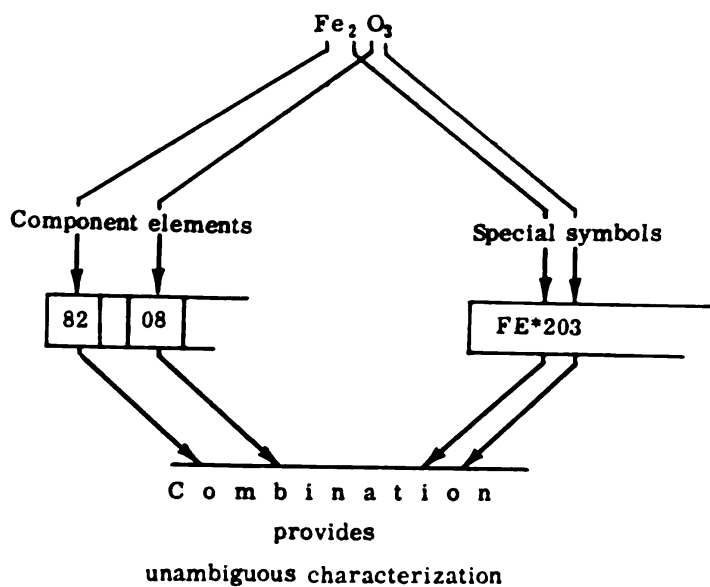
The citation of a given element combination, for example, the provision of a readable formula for a compound, must take into account the presently available capabilities of the punched card and of the various machines, such as the interpreter and the tabulator, that are operated by the cards. This means, however, that every formula must be set up as a linear array of symbols (indexes to indicate the number of atoms per molecule, exponents to indicate the charge on an ion, or to identify a given isotope). It must as a consequence be so interpreted that the linear character of the formula is maintained. The same applies also to punctuation marks, for example, the use of the period to indicate water of hydration in many formulas. This requirement of linearity of formula presentation made it necessary to provide a few symbols which have been found essential in dealing with substances occurring in inorganic chemistry, and in particular the chemistry of complex compounds. See Figure 28-6. It will be observed that the symbols as listed do not collide with the element symbols and that the formulas can be printed without danger of ambiguity arising. Examples of the application of these 12 symbols to a series of chemical compounds is shown in Figures 28-7 and 28-8.

A combination of letters and numbers are used for punching chemical formulas in the card in such a form as to permit formula listing. For this purpose, 31 columns of the punched card are reserved, namely, columns 50

* = Preceding letter is small	G = -
* = Preceding number remains on the line	M = .
E = (	R* = ,
Q =)	L = positive ion
X = [	LL = negative ion
Z =]	= Symbol for empty column

Number,	at the start or after an empty column, becomes a superscript	2H = ² H
	with asterisk, remains on the line	2*H = 2H
	without special designation, becomes a subscript	H2 = H ₂

Figure 28-6. Special symbols for encoding inorganic formulas in a linear form.



	Number field	Alpha field
Iron (element)	H0 1 00 00 00 00 00	Iron (element)
Iron oxides	H0 3 08 00 00 00 00	Iron oxides
Fe	82 1 00 00 00 00 00	FE*
FeO	82 3 08 00 00 00 00	FE*0
Fe ₂ O ₃	82 3 08 00 00 00 00	FE*203
Fe ₃ O ₄	82 3 08 00 00 00 00	FE*304

Figure 28-7. Examples of encoding of inorganic compounds.

$H_2^{32}SO_4$	H2 32SO4
$[SiF_6]^{2-}$ - Ion	XSI*F6Z 2LL ION
$KHCO_3 \cdot MgCO_3 \cdot 4H_2O$	KHCO3MMG*CO3M4*H2O
$H_2SO_4 - Na_2SO_4 - ZnSO_4 - H_2O$	H2SO4GNA*2SO4GZN*SO4GH2O
$(NH_4)_3H_5[Co(OH)(MoO_4)_3] \cdot 3H_2O$	ENH4Q3H5XCO*EOHQEMO*O4Q5ZM3*H2O

Figure 28-8. Further examples of encoding of inorganic compounds

to 80, inclusive. See Figure 28-9. This number of columns has proved sufficient for accommodating 95 per cent of all of the formulas encountered by Gmelin in the literature of inorganic chemistry and, in particular, in the chemistry of the complex compounds. In those few cases, when the length of the formula requires more than 31 columns, a trailer card is provided which is linked to the main card by an appropriate coupling symbol.

The card layout for chemical compounds might be summarized as follows:

<i>The Number Field Requires</i>	<i>Number of Columns</i>
for leading element	2
for the 5 following elements (5×2)	10
in all for the characterization of the elements present in a combination	12
to indicate the form or condition in which a material exists or is described	1
so that the number field (columns 2 to 14) in total includes	13

The column which indicates state or condition of a substance follows immediately after the two columns for registering the leading element and serves in an orienting capacity to indicate the nature of the material in question. A distinction is made between a single element and an element group. These general subdivisions divide the factual information pertaining to any one leading element into four main groups.

	<i>Element Group</i>	<i>Single Element</i>
Element	1	5
Alloy	2	6
Compound	3	7
Mineral	4	8

For the unambiguous characterization, e.g., of a compound, it is necessary to take into account the entries for this compound both in the number field and in the alpha field. The number field, as already explained, is essentially limited to indicating which elements appear in the compound. The formulation for tabulation of a formula in readable form in the alpha

81		76	29	08	07	00	ENH4Q3H5XCO•EOHQEMO•O4Q5ZM3•H2O
Number field							Alpha field
13 Columns							31 Columns

Figure 28-9. Fitting the code for an inorganic compound into the fields of an IBM card.

field provides the necessary information to characterize the compound as to stoichiometric proportions, ion charge, etc. See Figures 28-7, 28-8, 28-9. For the alpha field, as already mentioned, 31 columns are reserved and, in fact columns 50 to 80.

Machine discrimination, that involves, e.g., FeO and Fe₂O₃ and the requirement to select out Fe₂O₃, is possible with the help of the sorting machine by directing a selecting operation to the fourth column in the alpha field. See Figure 28-7. The collator can be used for this purpose when the comparison card is punched for the second position in the fourth column of the alpha field. Further features of the information relative to inorganic chemistry can be recorded in the card so as to take into account a large measure of data pertaining, for instance, to physical properties and to physical-chemical characteristics.

(a) For elements; form of occurrence 1 or 5 in columns 4 to 13 of the number field and more specifically citation of the modification in accordance with a special code in columns 4, 5, 6, 7, citation of valency in columns 8 and 9, and citation of isotopic forms in columns 10-13

(b) For alloys and compounds; form of occurrence 2 or 6 in the alpha field. After recording the formula in a form permitting readable tabulation, any remaining columns can be used for recording the trivial name or some similar citation.

For factual information pertaining to a given substance, the system as explained on pages 594 to 596, is applied: for the main divisions, principal group and subgroup, two columns each are used and for the specific subject, 3 columns, that is to say 9 columns in all. These are columns 17 to 25 in the machine-punched card.

For the literature citation the following scheme is applied:

	Column
Kind of publication (AB)*.....	1
Periodical, indicated by a specially assigned code number (z)*.....	4
Series (S)*.....	2
Volume (H)*.....	2
Serial set (R)*.....	1
Volume (Bd)*.....	4
Year (J)*.....	2
Numbers of pages (from — to —) (Seite v. b.)*.....	8
Total.....	24

* German abbreviations refer to Figure 28-11.

The reason for reserving 8 columns to indicate the pages occupied by a paper is because it is very advantageous, for example, when ordering a microfilm or photocopy, to know the actual length of a given paper.

For the citation of the literature (bibliographic citation) columns 26 to 49 on the machine-sorted punched card are reserved, as indicated below (See Figure 28-10 for punching instructions for recording a bibliographic citation).

Summary Review of the Card Layout

Column No.	Characterization	No. of Columns
1	Kind of card (KA)*	1
2, 3	Leading element (LE)*	2
4	Form of the substance (Art.)*	1
5-14	Following elements (Folge-Elem)*	10
15, 16	Condition (Zust)*	2
17, 18	Main division (Ü)*	9
19, 20	Main group (H)*	
21, 22	Subgroup (U)*	
23-25	Specific subject (EF)*	
26	Kind of publication (Original, abstract, list of titles, etc.) (AB)*	1
27-30	Periodical (Z)*	4
31, 32	Series (S)*	2
33, 34	Volume (H)*	2
35	Book series (R)*	1
36-39	Volume (Bd)*	4
40-41	Year (J)*	2
42-49	Numbers of pages	8
50-80	Readable formula of the substance, trivial name or the like (Leitstoff)*	31
		80

* German abbreviations and terminology refer to Figure 28-11.

Figure 28-11 shows the card layout corresponding to the summary given above. Figure 28-12 is an example of a card that was punched to record information pertaining to Fe_3O_4 as provided by a certain literature reference.

The minerals constitute a special case⁶² for which the automatic registration of information and formulas on punched cards has received particular attention. The starting point for dealing with minerals is, of course, their chemical formula or the formula of their components. After the leading element has been recorded in the same way as with chemical compounds, there is recorded for the minerals, in place of the following elements, a system for characterizing minerals as follows: class (Klasse), main group (H. Gr.), subgroup (U. Gr.), individual minerals (Min.), and, when necessary, various varieties (Var.) of individual minerals. (See Figure 28-13). Subdivision into classes corresponds in large measure to the usual mineralogical system as worked out, for example, by Strunz and Dana. This

Figure 28-10. Encoding of bibliographic references (examples).

Figure 28-11. IBM card layout for Gmelin documentation.

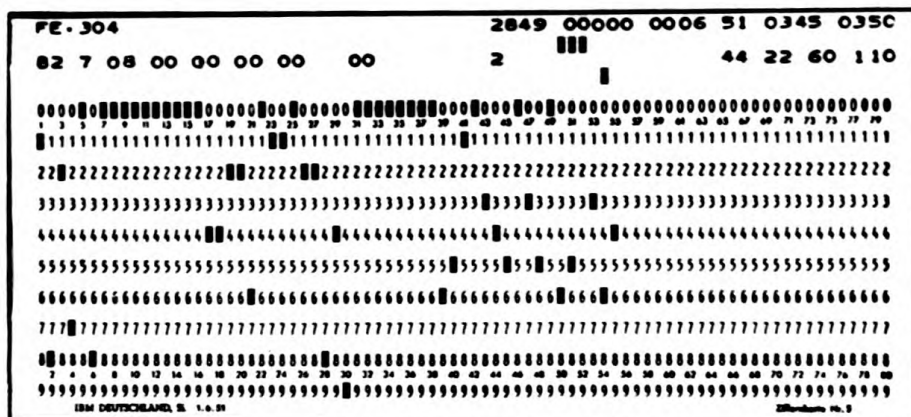


Figure 28-12. Sample card punched for information pertaining to F_3O_4 .

[illegible]

Figure 28-13. Card layout for minerals.

system groups together minerals with the same or similar anions. Without going into the details of this system for minerals, which has been presented elsewhere⁶², an example will be given to show how a specific mineral, namely, hydroxylapatite was entered in the Gmelin punched card. (See Figure 28-14). The Gmelin literature card records not only the class and subclass to which a given mineral pertains, but also indicates its component elements (enthaltene Elemente). See Figure 28-13. As a consequence it is possible to select out by automatic processes all minerals which contain a given element or a given combination of elements. This is accomplished by tabulating a list of the minerals in question which can be subgroups according to various points of view or characteristics. These procedures make it possible for the first time to compile lists of minerals occurring in nature on the basis of their component elements and to direct information searches to minerals containing specified components.

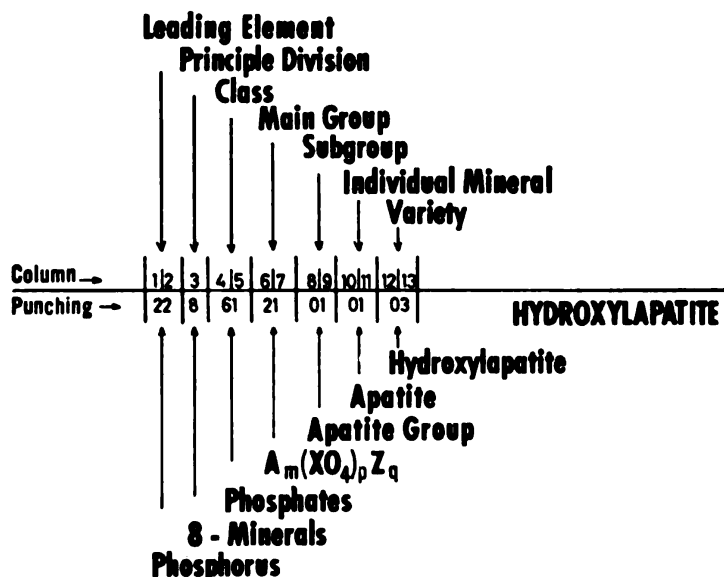


Figure 28-14. Encoding and punching for the mineral hydroxylapatite.

The Gmelin Literature Searching Service

The continually expanding volume of scientific and technical literature makes it increasingly impossible for any handbook to keep pace with new publications³⁶. This means that at any given point in time, only a relatively small number of volumes of the handbook are up-to-date, that is to say, present the state of science in the sense that Berzelius used this term. Most of the previously issued volumes of the handbook were published several years ago, while other volumes in the 8th edition of the handbook have not yet appeared. This means of course that the user of the handbook may not find the information that he is seeking in conveniently summarized form. In preparing the 8th edition, it was necessary, of course, to establish a literature closing date but this has meant that for many fields covered by the Gmelin handbook the time interval between the appearance of a volume devoted to a given element and the literature closing date may be considerable. Nevertheless, it is certainly desirable, important and advantageous to make it possible for the users of the Gmelin handbook to inform themselves concerning the state of knowledge on any given subject or any given area of specialization rapidly and thoroughly, in accordance with Gmelin traditions. To serve this purpose, a Gmelin literature searching service has been established. This service is based on the resources of the Gmelin documentation division and its archives which are maintained on a current basis. Requests for information and questions within Gmelin's broad field are to be serviced by supplying the literature which has appeared either between the previously published volume and the present time, or the total

literature for those volumes of the 8th edition which have not yet been published. In addition to answers to specific questions, the literature searching service provides its clients with compilations that are prepared at regular intervals to cite the literature of pertinent interest to various requested subjects or fields of specialization. In providing these literature services, the material is furnished in the form of archive cards which carry, by agreement with the American Chemical Society, the corresponding abstract of the paper in question from *Chemical Abstracts*. See Figure 28-4. In many cases, however, the Gmelin documentation center processes the original literature before abstracts are published and, in such cases, the Gmelin archive cards are furnished to the client. The archive cards are either German standard file cards (DIN-A-6 or DIN-A-5) or, at the client's request, hand-sorted punched cards (either DIN-A-5 or other requested format)⁶³. In furnishing information services in these ways, charges are computed so as to cover the costs incurred by the information service group within the Institute. In providing its services, this information service group makes use of the "classical archives" as well as the more recently generated card files that can be searched automatically.

The Application of Hand-Sorted Punched Cards in Gmelin Documentation

The Gmelin Institute has developed—as a German DIN standard—a form of hand-sorted punched cards⁶³ for which various special subject and author codes have been worked out and applied for particular purposes by the scientific staff.^{63a} In particular, when it is necessary to study and evaluate about 500 to 5,000 original papers in order to write a given section of the handbook, it becomes advantageous, even necessary, to be able to organize the numerous references on the basis of a multiplicity of subjects. Hand-sorted punched cards have been found to contribute greatly to the efficiency of the scientific staff members in utilizing such special literature collections and in writing manuscripts for the handbook. For simple cases, selection is accomplished with a sorting needle. For questions that involve a multiplicity of code entries, a sorting device developed at the Institute may be used⁶⁴.

The Application of "Peek-a-boo" Cards in Gmelin Documentation

The so-called "Peek-a-boo" cards provide a particular simple example of how mechanical aids may be used to analyze complex concepts into underlying basic concepts, that is to say, into simpler, independent underlying ideas⁶⁶. We will demonstrate this by considering the highly specialized field of platinum complex compounds which are characterized by exceptional diversity, complexity, and the occurrence of numerous isomers. This demonstration will be based on the "Sphinxo" cards as developed by R. Gagarin.

The visual selection of centrally punched cards as worked out by W. E. Batten and G. Cordonnier is the basis for this general method⁶⁶. It should be pointed out that a close analogy exists between the codification of a complex class of chemical compounds, in our example the platinum complex compounds, in terms of their components and the corresponding codification of a complex piece of information in terms of the underlying ideas and concepts. Thus, the same general mechanical methods as applied by "Peek-a-boo" cards have been found equally advantageous when dealing with complex compounds and with involved information statements.

As an example of how these cards are usefully applied to the special problems of Gmelin information processing, we will consider the compound (Pt a py₃) (PtCl₄) which has been assigned the compound serial number 440. The "analysis," that is to say, the codification of this compound, is carried out in the following fashion:

	<i>Empirical Code Catalog Number</i>
1. Type [Pt A ₃ B]X ₂ or [Pt AB ₃]X ₂	4
2. One coordinated component A.....	A1
3. Three coordinated components B.....	A13
4. Ammonia = a.....	102
5. Pyridine = py.....	142
6. Tetrachloroplatinate (II) ([PtCl ₄] ⁻² as anion X ₂).....	519

In all cases, as extensive experience extending over many years has proved, it is possible to arrive at the complex structures in an associative fashion from the component units (terms) so registered. In practical application, the Gmelin Institute uses the "Sphinxo" cards which are a further evolution of those originally developed by W. E. Batten and G. Cordonnier⁶⁷. Each subject heading is assigned to a given card just as in the "Uniterm" system. In each of these cards, there is then punched a number which indicates either a given substance or a given literature citation to which the characteristic indicated by the card pertains. If it now becomes necessary to determine which substances or which literature references have the characteristic in common, it is only necessary to consult the card to which the subject N has been assigned and to read off the punched numbers which correspond to the sought for substances or literature citations. If the search is directed to substances or literature citations characterized by two or more characteristics, for example, A, N, and Z, then one selects the cards to which A, N, and Z have been assigned and superimposes them. Those perforations which go through all three cards correspond to those items which are of interest for the search. See Figure 28-15. A "negative selection," that is to say, determining which substances or literature citations have characteristic A but do not have characteristic B can be carried out simultaneously with the determination of which items possess both characteristics

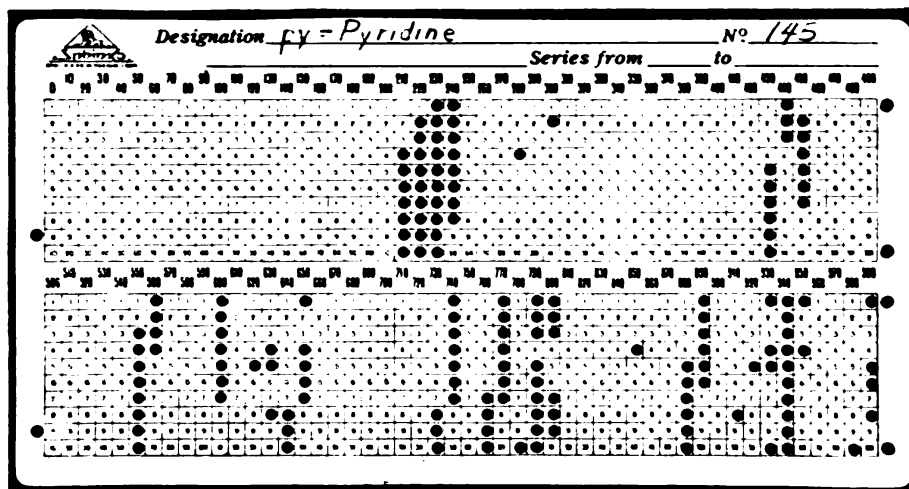


Figure 28-15. Example of "Sphinxo" card as used for complex compounds of platinum.

A and B. To this end one proceeds as follows: A colored transparent sheet is placed between the two cards when they are superimposed so that the color of the bottom card is thereby changed. Then the two cards with the interposed transparent colored foil are held up to the light, the negative selections correspond to opaque holes of a color which is different than that of the transparent colored holes corresponding to those punched for both A and B.

Over-all Organization of the Gmelin Institute

Previous paragraphs have discussed the tasks assigned to various groups within the Gmelin Institute. These groups have been organized to work together in such a fashion as to achieve the greatest possible over-all efficiency. The general organization chart, as given in Figure 28-16, makes it clear that all of the supporting divisions (for example, library, documentation division, phototechnical division, information service) work together with and support the scientific sections of the Institute and, in particular, the scientific and editorial staffs, that is to say, the fundamental and most important groups. It is perhaps clear that the supporting divisions accomplish a considerable variety of tasks which support the efforts and activity of the scientific staff. The latter's ability to perform assignments is of course the decisive factor determining the efficiency with which the Institute accomplishes its mission.

Automation of Gmelin Documentation—Future Trends

Although the form of mechanical documentation which Gmelin has worked out and has been applying in its present research effort appears

The flowchart illustrates the organizational structure and workflow of the Library and Documentation Division. At the top, 'Institute Management' oversees the 'Library' and 'Documentation division'. The 'Library' provides 'Literature required and supplied' (from Gmelin library and interlibrary loans) to the 'Scientific staff'. The 'Documentation division' handles 'Forwarded literature' and 'Preliminary clarification and interpretation'. It includes 'Abstract sources', 'Abstract cards', 'Information assignment', and 'Original periodicals'. 'Literature references' are processed into 'Duplicated cards' and 'Photographic division'. 'Classical card archives' and 'Abstract card archives' are maintained. 'Preparation and analysis' leads to 'Instructions for preparing punched cards'. 'Materials and subject contents' are processed into 'Bibliographic lists' and 'Abstract cards (copies)'. 'Coding' and 'Punching instructions' are used for 'Information service cards'. The 'Information service' provides 'Information service cards' to 'Information to clients' and 'Incoming questions'. The 'Editorial staff - Division chief' manages 'Manuscript subject assignment' and 'Manuscripts'.

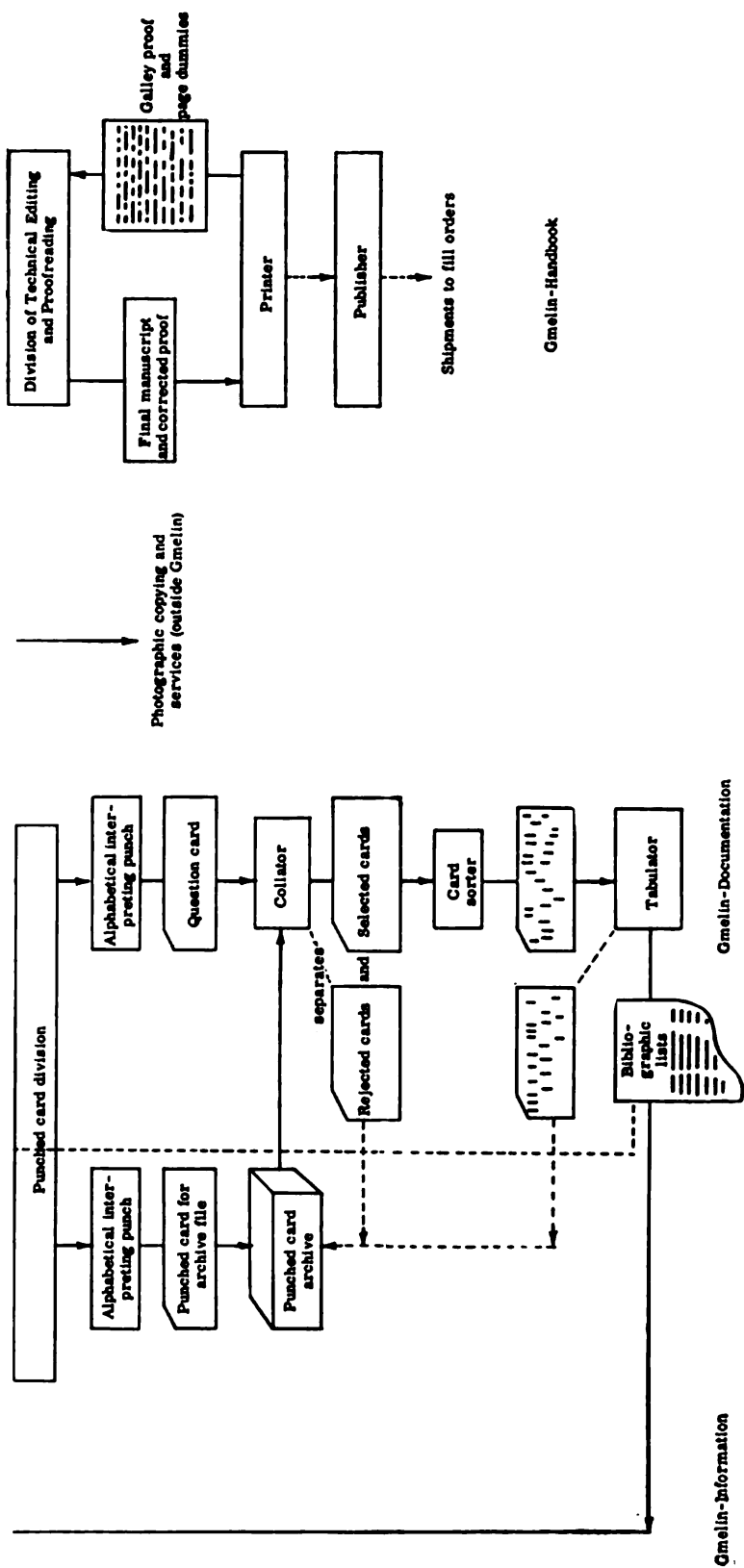


Figure 28-16. Organization chart of the Gmelin Institute.

capable of satisfying previously formulated requirements, it has nevertheless appeared advisable and necessary to consider carefully whether and to what degree, and at what point in time it may be advantageous to go over to purely electronic storage aggregates. In this connection, an estimate has been made of the annual amount of new information which probably will have to be stored. When this amount of information is expressed in terms of yes-no decisions, that is to say, in bits, then a comparison is possible with the present storage capacity of various types of electronic equipment.

At the present time (1958) the volume of documentary information produced per year within Gmelin's sphere of interest, that is to say; literature citations, subject headings and the text of the abstracts for about 40,000 original papers may reach the magnitude of 10^7 to 10^9 bits.

The storage capacity of magnetic memories as exemplified by the IBM "Ramac" is such that for each memory unit 10^6 bits could be stored⁶⁸. With the photoscopic units employing glass plates as developed by Telemeter, 10^8 bits per unit may be stored⁶⁹. It would appear, then, that the documentary information accumulating each year and of interest to Gmelin, could be accommodated without difficulty with the aid of storage units as they are provided now by electronic equipment manufacturers. The present state of electronic development permits us to conclude that the development of automatic documentation can be expected to move in the direction of electronic storage units.

References*

1. Jungk, R., "Die Zukunft hat schon begonnen," 11. Edition, Stuttgart 1954.
- 1a. Pietsch, E., "Was kann und was muss die Dokumentation für die Forschung und Technik leisten?" *Nachr. Dokumentation*, 9, 57 (1958).
2. Pietsch, E., "Dokumentation und Wissenschaft, Carl Bilfinger zum 75. Geburtstag," *Beiträge zum ausländischen Recht und Völkerrecht*, 29, *Völkerrechtliche und Staatsrechtliche Abhandlungen*, p. 322, Cologne, 1954.
3. Herberts, K., "10.000 Jahre Malerei und ihre Werkstoffe," p. 13, 2nd Edition, Wuppertal, 1938, p. 13; Pietsch, E., "Die Urgeschichte der chemischen Technologie und deren Bedeutung für die prähistorische Kultur," *Technikgeschichte*, 29, 15 (1940).
4. On the general subject of neolithic cultures and stone drawings, see Kühn, H., "Die Feldsbilder Europas," Stuttgart, 1952.
5. Erman, A., "Die Hieroglyphen," Sammlung Götschen, Nr. 608, Leipzig, 1917.
6. Otlet, P., "Traité de Documentation. Le livre sur le livre," Bruxelles 1934; Lafontaine, H., Otlet, P., "Création d'un répertoire bibliographique universel, note préliminaire," *Bulletin de l'Institut international de Bibliographie*, 1895/96.
7. Pietsch, E., Hantke, G., Mulert, G., "Massgebliche Dokumentationseinrichtungen in den Niederlanden," *Nachr. Dokumentation*, 5, 13 (1954).
8. Pietsch, E., "Grundfragen der Dokumentation," *Arbeitsgemeinschaft für Rationalisierung des Landes Nordrhein-Westfalen*, 14, 22 (1954).

* Compiled by staff librarian, G. Mulert.

9. Büchen, W., "6. Jahresversammlung der Deutschen Gesellschaft für Dokumentation in Bad Brückenau," *Nachr. Dokumentation*, 5, 219 (1954).
10. Hyslop, M. R., "Inventory of Methods and Devices for Analysis, Storage and Retrieval of Information," Chapter 6 in "Documentation in Action," ed. by J. H. Shera, A. Kent, J. W. Perry, Reinhold Publishing Corp., New York, 1956.
11. Pietsch, E., "Über Notwendigkeit, Wesen und Bedeutung der mechanisierten Dokumentation," *Manual on Document, Reproduction and Selection*, Section 711, Part II: Selection, Fédération Internationale de Documentation, The Hague, 1957.
12. Scheele, M., "Das Lochkartenverfahren in Forschung und Dokumentation mit besonderer Berücksichtigung der Biologie," Stuttgart, 1954, 114 pp.
13. Hosemann, H., "Das Lochkartensystem und einige wichtige statistische Resultate auf gynäkologisch-geburtshilflichem Gebiet," *Bibliotheca Gynäcologica* 10, 77 (1951); "Medizin und Dokumentation," *Nachr. Dokumentation*, 4, 152 (1953).
14. Fritze, E., "Die Auswertung von Krankenblättern mit Randlochkarten," *Nachr. Dokumentation*, 7, 134, (1956).
15. "Bericht über die Arbeitstagung des Ausschusses Medizin in der DBD zum Thema: Lochkarten und Randlochkarten im klinischen Gebrauch—am 4. und 5. Mai 1956 in Göttingen," *Nachr. Dokumentation*, 7, 230 (1956).
16. Pietsch, E., "Arbeitstagung des Ausschusses zur Mechanisierung der Dokumentation in der Deutschen Gesellschaft für Dokumentation in Berlin vom 4. bis 6. Dez. 1952," *Révue de la Documentation*, 19, 101 (1953).
17. Aristotle (384–322 B.C.) instigated the writing by Theophrastus of his work "*Φυσικῶν ἀδελφῶν*" on the early history of philosophical-cosmogonic thought. Aristotle was also responsible for Eudemos' writing the history of early astronomy and mathematics.
18. The "Historia naturalis" in 37 volumes of Caius Plinius Secundus (23–79 A.D.) was concerned with the origin of the universe, geography, ethnology, anthropology, and the natural exact sciences.
19. Martianus Capella's (before 429 A.D.) "De Nuptiis Philologiae et Mercurii", *Encyclopedia of the Seven Free Arts* (Quadrivium: Geometry, Arithmetic, Astronomy, Music; Trivium: Grammar, Rhetoric, Dialectic), translated into German by Notker III of St. Gallen in 1022.
20. A.M.S. Boethius (480–524 A. D.) translated, with commentary, numerous philosophical and mathematical works of the Greeks, and so played an important role in making ancient knowledge available in the Middle Ages.
21. Isidore of Seville wrote his extensive "Originum sive Etymologiarum libri xx" in 622–633.
22. The work of monastic orders, especially of the time of Charlemagne, played a decisive role in the transmission of ancient knowledge during the fifth and sixth centuries. Compare also: J. M. Clark, "The Abbey of St. Gall as a Centre of Literature and Art," Cambridge, 1926, also, Reference 24.
23. Bartholomaeus Anglicus' (thirteenth century) "De proprietatibus rerum", written between 1220 and 1230, embraces all the knowledge of the time, with the exception of history. This work was based on that of Isidor of Sevilla (570–636 A.D.), whose twenty-volume compilation, "Etymologiae", is a first-rate source book (quoted in 660 chapters by Bartholomaeus Anglicus), particularly for the natural sciences based on Aristotle (quoted in more than 275 chapters from Arabic translations). Consideration was given to over one-hundred ancient authors. The work of Bartholomaeus Anglicus exerted a profound influence for

- three centuries on European intellectual development. Douce calls him "our English Pliny". Shakespeare used Anglicus' work as an authority for astronomical questions; Dante used it when preparing the Italian translation of Belcalzer, "Trattato di Scienza universal" (1307 or 1308). Compare: Se Boyar, Bartholomaeus Anglicus and his "Encyclopedia" in *Journal of English and German Philology*, 19, 168-89 (1920).
24. Thomas de Chantimpré's (1201-1263 or 1280) "Liber de natura rerum," a work of nineteen volumes published in 1232 after about fifteen years preparation. At that time the author wrote "For all of fourteen or fifteen years we have devoted ourselves most diligently to books on the nature of things, always keeping before our eyes the sentence from the book on the Christian doctrine by St. Augustine (354-430 A.D.) that it would be useful to the highest degree if someone would undertake the task of compiling in one volume the nature of things and especially that of living creatures". In 1472 it was translated into German by Peter Königschlafer, school teacher and city scribe in Waldsee. In 1478 Michael Bauman published his twenty-volume encyclopedia on natural history, basing it mainly on Thomas. Compare A. Kaufmann's "Görres-Gesellschaft zur Pflege der Wissenschaft im katholischen Deutschland", Erste Vereinsschrift für 1899, Cologne, 1899.
 25. Vincent de Beauvais (Bellovacensis) (died about 1264) wrote his "Encyclopedia Speculum Maius (Triplex)" at the instigation of the French King Louis IX, and added a "Speculum morale (Speculum quadruplex)" for the fields of natural sciences, theology, history, and philosophy. In many places it is based on Thomas de Chantimpré, without naming him. Compare: "Vincent de Beauvais et la connaissance de l'antiquité classique au treizième siècle," *Revue des questions historiques*, IXème année, 17, 5 ff, Paris (1875).
 26. Although based on Thomas of Chantimpré, Konrad von Megenberg's (1309-1374) "Das Buch der Natur" has numerous individual observations and critical notes on the earlier authors. Von Megenberg's work appeared in 7 volumes around 1350. Its sixth volume deals with precious stones, the seventh with metals. But, characteristically for the state of knowledge, his discussion of the seven metals, when printed, did not occupy more than 50 lines. Compare the edition: "Das Buch der Natur" von Konrad von Megenberg; Die erste Naturgeschichte in deutscher Sprache; In neu-hochdeutscher Sprache bearbeitet und mit Ammerkungen versehen," (edited in modern German with comments by Dr. Hugo Schulz, Greifswald, 1897.)
 27. "Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers," Mis en ordre par Diderot et par d'Alembert. Tomes 1-17, Planches, 1-11, et Supplementes 1-5. Paris, 1751-80.
 28. We shall here dispense with an enumeration of individual publications. The extent of G. W. von Leibniz's profound knowledge encompassed widely diverse fields. Thus, in mathematics he was the founder of differential calculus. He was also active in the natural sciences. In philosophy his work in monadism might be cited. He was called an "academy" of science and in fact successfully concerned himself with the founding of scientific academics in Berlin and St. Petersburg. His complete works must be consulted in order to comprehend his accomplishments fully.
 29. Alexander von Humboldt, "Kosmos. Entwurf einer physischen Weltbeschreibung," 4 Bände. Stuttgart, 1845, or 1862. See statements in E. Pietsch "Sinn und Aufgabe der Geschichte der Chemie," *Angew. Chemie* 50, 939 (1937).
 30. Pietsch, E., "Johann Rudolph Glauber—Der Mensch, sein Werk und seine Zeit,"

Deutsches Museum, Abhandlungen und Berichte, München: Oldenbourg, 1956, 64 pp.

31. Chemistry and medicine were very closely associated from the first quarter of the sixteenth century to the middle of the seventeenth century. Paracelsus' (1493–1541) contributions came in the middle of this period, uniting the two fields by his earlier work. Boyle finally gave chemistry its distinctive character. For works of pronounced medical nature consult: J. B. van Helmont (1577–1644), "*Ortus medicinae vel opera et opuscula omnia*," Amsterdam, 1648 and its English, French, and German translations.
32. A. Libavius (1540 or 1550* until 1616) wrote the first handbook of chemistry: "*Alchymia*," 1595.
33. For the seventeenth century the following appear particularly worthy of mention here: N. Lefébvre (= Le Fébre), "*Traité de la Chymie*," Paris, 1660, which rapidly passed through 5 editions; C. Glaser, "*Traité de Chymie*," Paris, 1663; N. Lemery "*Cours de chymie*," Paris, 1675 (for a long time a leading text for chemists and pharmacists); J. H. Jüngken, "*Lexicon chemicopharmaceuticum in duos partes*," Nürnberg, 1699. For the eighteenth century, besides H. Boerhave, "*Elementa Chemiae*," Leiden-London-Leipzig-Paris, 1732, and the numerous works of Becher and Stahl (on this compare H. Kopp, "*Geschichte der Chemie*," Vol. 1, Braunschweig, 1843), let us mention first of all the textually clear six-volume "*Dictionnaire de chimie*" by Macquer, which was published in 1767. German edition: D. J. Macquer, "*Chymisches Wörterbuch oder allgemeine Begriffe der Chymie . . . mit Anmerkungen und Zusätzen vermehrt von J. G. Leonhardi*," Leipzig, 1781–1791. Other important works include: Wiegand, "*Handbuch der allgemeinen Chemie*," Berlin, 1781, English edition in 1789; A. L. Lavoisier, "*Traité élémentaire de chimie présenté dans un ordre nouveau et d'après les découvertes modernes*," Paris, 1789. The latter heralded the modern era of chemistry. The new nomenclature used in this work was drafted in 1782 by Guyton de Morveau, Bernard, Lavoisier, Berthollet, and Fourcroy. At the end of the eighteenth century we find the three-volume "*Geschichte der Chemie*" by Johann Friedrich Gmelin, Göttingen, 1797 to 1799.
34. The sequence of publication of the "*Gmelin Handbuch*" might be summarized as follows: "*Handbuch der theoretischen Chemie*," first edition, three volumes, published by Varrentrapp, Frankfurt, 1817–1819; Second edition, published by Varrentrapp, Frankfurt, 1821–1822; third edition, published by Varrentrapp, Frankfurt, 1827–1829; and fourth edition, published by the Universitätsbuchhandlung of Carl Winter, Heidelberg. It started from 1843 as "*Handbuch der Chemie*." Volumes 1 to 4 were still edited by L. Gmelin; from volume 5 on edited by Kraut and List, according to his (Gmelin's) manuscripts. The fourth edition was published in English by the Cavendish Society in 1846 (with supplements by Watts). Fifth edition, published by the Universitätsbuchhandlung of Carl Winter, Heidelberg, from 1852, as "*Handbuch der anorganischen Chemie*." This edition was started during Gmelin's lifetime, but was edited by Kraut and List. Sixth edition, published by the Universitätsbuchhandlung of Carl Winter, Heidelberg, from 1877. Seventh edition, published by the Universitätsbuchhandlung of Carl Winter, Heidelberg, from 1907. Eighth edition, published as the "*Gmelin Handbuch der anorganischen Chemie*," *Verlag Chemie*, Berlin, from 1924. Published by the Deutsche Chemische Gesellschaft in *Verlag Chemie*. From 1945–1950, published by the Gmelin-Institut in the Gmelin-Ver-

* According to more recent investigations by Dr. Atterer of the Gmelin Institut it is improbable that Libavius was born before 1550.

- lag G.m.b.H. From 1950, published again by *Verlag Chemie*, Weinheim. Edition started by R. J. Meyer, and continued by E. H. E. Pietsch.
35. Compare E. Pietsch, "Bericht über das Gmelin-Institut für anorganische Chemie und Grenzgebiete in der Kaiser-Wilhelm-Gesellschaft zur Förderung der Wissenschaften," (Clausthal-Zellerfeld), *Experientia*, 3, 8 (1947); H. B. Hass, "Present Status of Beilstein and Gmelin," *Chem. Eng. News*, 26, 1430 (1948); and A. R. Todd, "International Union of Chemistry's Advisory Council on Beilstein and Gmelin," *Ibid.*, 3621; Pietsch, E., *J. Chem. Educ.*, 26, 251-253 (1949); Pietsch, E., "Aus der Arbeit am Gmelin-Handbuch der anorganischen Chemie zum hundertsten Todestage von Leopold Gmelin," 13. April 1853," *Chimia*, 7, 49 (1953). Regarding Leopold Gmelin compare E. Pietsch and E. Beyer. "Leopold Gmelin der Mensch, sein Werk und seine Zeit," *Ber. deutschen chem. Ges.*, 72 (A), 5 (1939); also E. Pietsch, "Erinnerungsschrift aus Anlass der 150 Wiederkehr des Geburtstages von Leopold Gmelin," Berlin, 1938.
 36. Pietsch, E., "Das Gmelin-Institut für anorganische Chemie in der Max-Planck-Gesellschaft," *Achema-Jahrbuch*, 1953/55, p. 116 and *Achema-Jahrbuch* 1956/58 p. 245.
 37. Pietsch, E., "Ernst Telschow und der Wiederaufbau des Gmelin-Instituts für anorganische Chemie und Grenzgebiete, Aus der deutschen Forschung der letzten Dezennien." *Festschrift Dr. Ernst Telschow zum 65. Geburtstag*, Stuttgart, 1956, p. 308.
 38. Regarding the "Beilstein Handbuch" compare F. Richter, "Friedrich Beilstein, Gedanken zur hundersten Wiederkehr seines Geburtstages," *Angewandte Chemie*, 51, 101 (1938); H. B. Hass, "Present Status of Beilstein and Gmelin," *Chem. Eng. News*, 26, 1430 (1948); A. R. Todd "International Union of Chemistry's Advisory Council on Beilstein and Gmelin," *Chem. Eng. News*, 26, 3621 (1948).
 39. Compare also M. Pflücke, "100 Jahre Chemisches Zentralblatt," *Ber. deutschen chem. Ges.*, 62, 3132 (1929); and H. Harff, "Die Entwicklung der deutschen chemischen Fachzeitschrift," Berlin, 1941.
 40. Crane, E. J., "Chemical Abstracts," *Chem. Eng. News*, 35, 74 (1957).
 41. Bradford, S. C., "Documentation," Public Affairs Press, Washington, D. C. 1950.
 42. Pflücke, M., "Dokumentation, zugleich ein Bericht über den Weltkongress der Dokumentation vom 16-21 August, 1937, in Paris," *Angewandte Chemie*, 50, 955 (1937).
 43. Pietsch, E., "Grundfragen der Dokumentation," *Arbeitsgemeinschaft für Rationalisierung des Landes Nordrhein-Westfalen*, 14, 18, (1954).
 44. The field of biology may be cited as an example: *Monthly Bulletin on Scientific Information and Terminology*, (edited by UNESCO) 1957, No. 6, p. 9.
 45. Crane, E. J., "Chemical Abstracts Wartime Obstacle Race," *Chem. Eng. News*, 23, 1757 (1945); "Five Miles of Index," *Ibid.*, 25, 1188 (1947); "CA Today, the Production of an Abstract Journal," *Ibid.*, 26, 2190 (1948). Lewis, Ch. M. "The role of the professional society," Chapter 9 in "Documentation in Action," ed. by J. H. Shera, A. Kent, J. W. Perry, Reinhold Publishing Corp., New York, 1956.
 46. Pietsch, E., "Zur Frage nach dem sowjetischen wissenschaftlichen und technischen Schrifttum und seiner Beschaffung," *Seventh Meeting of the AGARD Documentation Committee*, Brussels, 30. August 1956 1/19 Ms.; Reichardt, G., "Sowjetische Literatur der Wissenschaft und Technik," Wiesbaden, 1957, 181 pp.
 47. "Report to the Secretary of Commerce by the Advisory Committee on Application of Machines to Patent Office Operations," U. S. Dept. of Commerce, Wash-

- ington, 1954, 76 pp. p. 3; Andrews, D. D., "Cooperative Information Processing-Patents," Chapter in "Documentation in Action," ed. by J. H. Spera, A. Kent, J. W. Perry, Reinhold Publishing Corp., New York, 1956, Staff Report "Patent Office Gets ILAS," *Chem. Eng. News*, 35 (1957) No. 17, p. 138.
48. Pietsch, E., "Dokumentation und mechanisches Gedächtnis. Zur Frage der Ökonomie der geistigen Arbeit," *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen*, 38, 33 (1954).
49. See, in this connection, E. Scott, in: "Mechanized system launches new era for literature searching. New IBM System will require language engineering to exploit its potentialities. Ultra-high speed electronic scanners predicted," *Chem. Eng. News*, 30, 2806, (1952).
50. Pietsch, E., "Dokumentation und Wissenschaft, Beiträge zum öffentlichen Recht und Privatrecht, Carl Bilfinger zum 75. Geburtstag gewidmet," *Völkerrechtliche und Staatsrechtliche Abhandlungen*, 29, (1954) 330.
51. Pietsch, E., "Dokumentation und mechanisches Gedächtnis. Zur Frage der Ökonomie der geistigen Arbeit," *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen*, 38, 44 (1954).
52. Pietsch, E., "Zur Frage der Kodifikation in der mechanischen Dokumentation," *Nachr. Dokumentation*, 7, 117, 179, (1956).
53. R. R. Shaw, "The Rapid Selector," *J. Documentation*, 5, 164 (1949/50) ; "Mechanical and electronic aids for bibliography," *Library trends*, 2, 522, (1954).
54. Samain, J., "Filmorex. Une nouvelle technique de classement et de sélection des documents et des informations," Paris 1952; Brochure, 1956, 11 pp.
55. Tyler, A. W., Myers, W. L., Kuipers, J. W., "The application of the Kodak Mini-card System to problems of documentation," *Am. Documentation*, 6, 18, (1955).
56. Bagley, P. R., "Electronic digital machines for High-Speed information searching," *Thesis presented to the Department of Electrical Engineering of MIT* (1951); Bagley, P. R., Perry, J. W., "Applicability of newer electronic techniques to information searching" 1952, 12 pp.
57. Cherry, C.: "On Human Communication," New York 1957; XIV, 333 pp. Goldman, S.: "Information Theory," London, 1953, XIII, 385 pp. Wiener, N.: "Mensch und Menschmaschine (The human use of human beings)," transl. Berlin 1952; L. Couffignal, *Les machines à penser* (translated by E. Walther: Denkmaschinen) Stuttgart, 1955, 186 pp.
58. For facts and figures concerning methods for producing the handbook, its scope and also for a bibliographic summary of the previously published sections of the Handbook, see "Facts about the Gmelin Handbook for Inorganic Chemistry, 8th edition", (Brochure, 1957).
59. "Gmelins Handbuch der anorganischen Chemie," ed. by the Gmelin Institute, Frankfurt am Main. "Systematik der Sachverhalte," Weinheim, 1957. VII, 116 pp. In German and in English. Vol. 2, "Alphabetic Listing" appeared in 1958. This subject contents summary was preceded by "Systematisches Stichwörterverzeichnis. Gmelin Institut Teil 1. Systematische Folge. Teil 2. Alphabetische Folge" (In German and in English) Clausthal-Zellerfeld, March 1, 1953. Staff members, who participated in developing the Gmelin key system, included: Dr. G. Hantke, W. Heimerdinger, R. Körber, K. Schneider.
60. Pietsch, E.:
(a) "Neue Methoden zur Erfassung des exakten Wissens in Naturwissenschaft und Technik," *Nachr. für Dokumentation*, 2, 38 (1951); (b) "Wie ist eine große Dokumentationsstelle für die Fachgebiete der Naturwissenschaften aufzubauen?" *Nachr. für Dokumentation*, 2, 116 (1951); (c) "Mechanisierte Doku-

- mentation—ein Weg zur Ökonomie geistigen Schaffens," *Umschau*, 52, 513 (1952); (d) "Aus der Arbeit am Gmelin-Handbuch der anorganischen Chemie zum hundertsten Todestage von Leopold Gmelin," 13. April 1953, *Chimia*, 7, 49 (1953); (e) "Wie bewältigt die anorganische Chemie ihr Schrifttum? Entstehung, Struktur und Bilanzierung des Gmelin-Instituts," *Chemische Industrie*, 5, 462 (1953); (f) "Mechanische Dokumentationsverfahren in der Praxis der Industrie," *Nachr. für Dokumentation*, 4, 82 (1953); (g) "Dokumentation und mechanisches Gedächtnis: Zur Frage der Ökonomie der geistigen Arbeit," *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen, Naturwiss. Reihe*, 38, 33 (1954); (h) "Methods of mechanical documentation," *International Congress on Documentation*, London, 1955.
61. Pietsch, E., "Die IBM-Karte in der Gmelin-Dokumentation," *Manual on Document, Reproduction and Selection*, Den Haag 1957 (Loseblattsammlung) Chapter 746.111.
 62. Kubach, I., (in Zusammenarbeit mit W. Müller) "Klassifikation der Mineralien für die automatische Dokumentation," *Fortschritte der Mineralogie*, 35, 38, (1957).
 63. Pietsch, E., "Dokumentation und mechanisches Gedächtnis. Zur Frage der Ökonomie der geistigen Arbeit," *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen* 38, 51 (1954); Pietsch, E., "Das Gmelin-Institut für anorganische Chemie in der Max-Planck-Gesellschaft," *Achema-Jahrbuch*, 1956/58, p. 245.
 - 63a. Pietsch, E., and Mulert, G., "Bibliographische Hilfsmittel der Chemie-Zur Anwendung der Randlochkarte in der Bibliographie," *Nachr. Dokumentation*, 6, 60 (1955).
 64. Ruston, W. R., "Die Randlochkarte als Hilfsmittel für die wissenschaftliche Dokumentation," *Nachr. Dokumentation*, 3, 5, (1952); Marginal punched cards, *Manual on Document Reproduction and Selection*, Den Haag 1957 (Loseblattsammlung) Kapitel 744.23.
 65. Pietsch, E., "Zur Frage der Kodifikation in der mechanischen Dokumentation," *Nachr. Dokumentation*, 7, (1956) 179; Gagarin, R., "Die Sphinxo-Sichtlochkartei für ein Teilgebiet der anorganischen Chemie," *Nachr. Dokumentation*, 8, (1957) 148;
 66. Holmstrom, E., "Superimposable cards," *Manual on Document Reproduction and Selection*, Fédération Internationale de Documentation, Den Haag, 1957 (Loseblattsammlung) Chapter 744.43.
 67. Kreithen, A., Gull, C. D., Miller, E. E., "Klassifizierung nach dem System gleichwertiger Grundbegriffe" (ins Deutsche übertragen von E. Pietsch und G. Mulert), *Nachr. Dokumentation*, 5, 6 (1954); "The Uniterm System of Coordinate Indexing," *Manual on Document, Reproduction and Selection*; Den Haag, 1957, (Loseblattsammlung), Kapitel 724.41; Taube, M., "Storage and Retrieval of Information by means of the Association of Ideas," *Manual on Document Reproduction and Selection*, (Loseblattsammlung), Den Haag 1957, Kapitel 724.42; Steier, H. P., "Uniterm: New Logic for Scientific Research," *American Aviation*, 27, August, 1956;
 68. IBM 305 Ramac, Random Access Memory Accounting, Manual of Operation: Preliminary Addition", pp. 81; Eckert, W. J., Jones R., "Schneller, schneller" ("Faster, faster:" ins Deutsche übersetzt von E. Aikele), Sindelfingen 1956 176
 69. King, G. W., "A New Approach to Information Storage," *Control Eng.*, August 1955.

Chapter 29

COMMERCIALLY AVAILABLE DATA PROCESSING MACHINES FOR FILING AND RETRIEVING TECHNICAL INFORMATION

ASCHER OPLER

*The Dow Chemical Company, Western Division, Pittsburg, California**

In the past few years, we have witnessed the rise of equipment designed to mechanize the handling of business files. In the first half of the twentieth century, punched-card techniques arose and took a dominating position in mechanized accounting and file processing. These machines are now beginning to give way to faster, more complex and more versatile electronic devices. In this chapter, we shall examine these machines and attempt to gauge their usefulness and significance when they are adapted to serve scientists in the same manner that hand- and machine-sorted cards are now used.

There has been a most productive cooperation between the commercial and the technological world. The punched-card equipment, for example, was developed for the use of machine accounting. In the late 1920's and in the 1930's, a handful of scientists began to make use of them in their own work. During the 1940's, when mathematical computations began to pervade technology, these machines came into widespread scientific and engineering use. At this point, this same scientific group developed practical electronic computers for their purposes and then, their use of the punched-card machines began to decline. As these electronic computers became reliable and efficient, the commercial users of punched-card equipment began to examine the prospects of using the fast computers as accounting tools in place of the mechanical punched-card equipment. Not only did the prospects look excellent, but the demand for this equipment proved so widespread that the use of computers in business applications has now outstripped the original scientific uses. Moreover, special computers (generally called data processing machines) were developed for commercial work, and techniques were invented for making better use of scientific-type computers for accounting purposes. Looking back, one can see the fruitful interplay of the requirements and the tools of both the technological world and the commercial world.

In many respects, the requirements for filing and subsequently retrieving

* Present address: Computer Usage Company, New York, N. Y.

business records (invoices, cancelled checks, warehouse receipts and withdrawals, etc.) are identical to the requirements for filing and retrieving scientific information. Both the business and scientific world deal with large masses of information that must be maintained and referred to with stringent accuracy requirements. While in the past there have been many proposals to build special machines for purposes of recovering scientific information, these proposals have not often been carried to realization because the technical complexity and the resultant cost have overbalanced the added value to be gained by automatic retrieval. Now that hundreds of business data processing machines are being distributed throughout the world, it appears that the need for such special machines will be reduced or eliminated.

Data Processing Machines: Description

A lengthy technical description of data processing machines is beyond the scope of this chapter. The reader will find such descriptions in the Bibliography. The manufacturers of such equipment also provide clearly written, descriptive brochures on the equipment.

The five classical parts of a computer or data processing machine are: (1) the input devices, (2) the output devices, (3) the storage or memory units, (4) the arithmetical or mathematical section, and (5) the central control unit that organizes and sequences the operation of the above under the control of a series of commands (program) stored within the computer's memory.

The following description of each of these parts is directed especially toward those interested in filing and retrieving technical information. The design of computing machines has undergone exceedingly rapid change in recent years and it is expected that this dynamic state will continue. Consequently, the material given below may become outdated before too long.

(1) **Input Devices** (carry information into the computer's storage unit in a form to be manipulated by the computer program).

(a) **Punched Cards.** Most data processing machines permit the use of punched cards for entering data and instructions. This entry may be direct or via intermediate magnetic tape. (See Section 1 (c) below.) The provision for entering data from punched cards stems from the long standing use of these cards with the resultant vast pool of information available on such cards and from the ease with which they may be punched and manipulated. It is still quite common to write the instructions which comprise the program on punched cards. This makes the frequent insertion, correction, and deletion required quite simple. Experienced users of data processors generally convert frequently used data and programs to magnetic tape and thenceforth dispense with the punched cards.

(b) **Perforated Paper Tape.** This mode of reading information into a computer is reserved primarily for programs on certain small machines and, in general, is of little interest to those in the information retrieval field. The input speed using tape is generally slow.

(c) **Magnetic Tape.** This medium, which may use either a plastic or metal base, serves in three of the computer functions, namely, input, output and memory. At the present writing, information stored on magnetic tape is undisputed first choice in the data processing field. The magnetic tapes themselves are relatively cheap, although the tape transport units for reading, writing, referencing and rewinding tapes under computer commands are not. Each standard length tape reel will hold roughly between one and fifty million alphanumeric characters. Information is stored as closely packed, magnetized and unmagnetized spots. Most tape reading and writing mechanisms provide for a redundancy check to detect random errors in the patterns representing the coded characters. The rate of read-in from magnetic tape will run from 1000 to 30,000 characters per second depending on the tape pattern used and the computer's organization.

(2) **Output Devices** (Write or punch information requested in a form which can be interpreted by humans. This information has been sent to the output device from the storage unit, under the control of the program.)

(a) **Punched Cards.** Many data processing machines have the ability to punch cards directly or indirectly via magnetic tape. This is often a desirable feature since the punched cards so produced may be compatible with existing files that are normally handled on punched card equipment. In some cases, punched card output is preferred because of the flexibility in format made possible by card manipulation and printing with tabulating machines.

(b) **Magnetic Tape.** (See Section 1 (c).)

(c) **Printed Output.** Printed reports and listings are available as direct output of many data processing machines, while others can produce results only through intermediate punched cards or magnetic tape. Since most tabulating devices in use now operate at a basic speed considerably slower than the data processing machine, the preferred method (especially for large output) is via magnetic tape. However, high-speed output devices are becoming increasingly available and recent demonstrations have been made of equipment printing 500, 600, 900 and 1000 lines per minute. In certain cases these can operate directly as "on-line" output of the data processing machine. Furthermore, a number of still faster devices are being developed but, thus far, none of these produce hard copy (ink on paper with multiple carbon copies produced simultaneously.)

(d) **Graphic Output.** A fourth class of output available on some machines provides for a two-dimensional graphical display. This may be

either in the form of charts (ink on paper) or, on some of the more sophisticated machines, the output may be displayed (and photographed) on the screen of a cathode ray tube.

(3) **Storage Devices** (Contain the program, sub-routines, working areas, and constants needed by the program to carry out the logical manipulation required. Also, they may contain the input information to be processed and may hold the output resulting from a calculation.)

(a) **Fast Storage Devices.** (1) *Magnetic Cores.* At the present time, the computers with fast storage are dominated by the magnetic core. Because of certain remarkable physical characteristics, they are capable of existing in only two magnetic states, which fits in well with the essentially binary logic required by a computing or data processing device. This bi-stable material is capable of being sensed or changed in approximately one millionth of a second. In addition to high speed, they are fast earning a reputation for astonishingly high accuracy. Memories assembled from over 100,000 of these cores have operated continuously for more than six weeks (3.6 million seconds) without detectable error.

The chief disadvantage of cores lies in the high cost of fabrication, testing, assembly and final testing. The mass production of the cores is done by powder metallurgy after material with suitable magnetic properties has been prepared, purified and reduced to powder. After fabrication, the hysteresis loop of each individual core must be checked and the unsatisfactory ones rejected. The most difficult task of all is that of assembly, since three wires must pass through each core and be fastened according to a rigid plan. To pass final inspection, every assemblage must be perfect: each core possessing a perfect hysteresis loop must be perfectly connected. Any single error of selection or fabrication makes the entire assembly unsuitable for use. Naturally, rigorous final testing procedures must be carried out on all these assemblages. It is primarily these mechanical and practical difficulties that prevent the more widespread use of cores.

At present core memories are in use in at least 100 large machines, in sizes from 70,000 to 300,000 cores. Plans have been announced for delivery of storage units containing over one million cores.

(2) *Williams Tubes.* These are cathode ray devices that were developed by Dr. Williams of Manchester University. They were the favored means of storage a few years ago. While many machines now in existence have this type of memory, there are none currently being produced that use it.

(3) *Mercury Column.* This was the first high-speed storage device used in commercial instruments. Like the Williams tube, this is no longer being produced, although many large computers still contain this device.

(b) **Intermediate Speed Devices.** (1) *Magnetic Drum.* For the past five years, the magnetic drum has held the undisputed leadership in computer storage. At least eleven commercial computer models have been built

entirely around magnetic drum storage and operation, while nearly every large scale computer employs auxiliary magnetic drums for secondary memories.

The advantages of the magnetic drum are obvious. Compared to the fabrication of magnetic cores, manufacturing magnetic drums is simple. A well machined, well balanced cylindrical rotor is coated with magnetic iron oxide. This is driven at moderate to high speed and read-write heads are fastened around the periphery. The surface area may be considered as sliced into narrow bands, each the width of one read-write head. A permanently recorded timing track appears on the drum and we may read or write any specified portion of the drum by reference to the timing track and the location of the band. This permits simple circuitry for storing and retrieving drum information.

The two disadvantages inherent in the magnetic drum are the relatively slow speed of access and a host of difficulties arising from the sequential flow of information past each read-write head. Because only a fraction (between 1/100th or 1/16th) of the information is under the reading head at any one instant, the machine must wait until the segment addressed appears under the head. In general, with completely random access, one must wait one half of one drum revolution, and thus somewhere around 49/50ths or 7/8ths of the time must be spent in waiting.

A number of schemes have been devised to circumvent this, including:

- (a) Non-sequential numbering of consecutive locations.
- (b) A special command address which directs the machine to a non-consecutive address for the succeeding command.
- (c) The use of special (fast access) bands which contain more than one read-write head. As the number of these heads increase, the waiting time is cut down proportionately.

Even when the above schemes are available for use and when they are used judiciously, the magnetic drum does not approach the speed or convenience of the magnetic core memory. However, it has proved to be the work horse memory of more than 1000 computers.

(c) **Slow Speed Storage.** (1) *Magnetic Tape.* These have been discussed in general under the sections on input and output. In those sections, their use is described mainly in association with peripheral equipment. Their use as storage falls into an additional category. Here the tapes may be used as "back-up memory" to enable the loading or unloading of the primary memory (usually magnetic cores or drums).

(2) *Newer Devices.* At present magnetic tape units are the accepted standard for slow access memory. There are a number of newer devices which have been developed and are soon to be marketed. No detailed account will be given here since operating experience has been negligible.

- (a) *Large, Slow Speed Magnetic Drum.* The usual magnetic drum will

store less than 4,000 words of information but these will generally be accessible at least 25 times a second. Recently drums are being considered that hold 25 times the information but make it available only once a second.

(b) *Photoscopic Storage Devices*. While photographic means of storage have been considered for some years, it is the development of this ingenious device which makes such techniques practical for fast machines. This unit consists of a transparent glass disk coated with photographic emulsion and which is photographically imprinted with a pattern of tracks containing opaque and transparent blocks. As the disk spins, the micro-blocks are scanned by a flying-spot scanner similar to devices used for transmitting slides by television. By the combination of the rotation of the disk and the sweep of the scanning tube, *all* spots are scanned each revolution. A single disk can hold a many as 15 million bits of information.

The inherent disadvantage for general purpose use is the non-erasable character of the stored information. Thus, it seems excellent where permanent information must be referred to (as in a dictionary or collection of information not subject to frequent revision) and should most probably be used in association with magnetic tape or other easily erasable storage.

(c) *The "RAM"*. This consists of 50 disks, each 24 inches in diameter and coated with magnetic material on both sides. Approximately 30 million bits may be stored on the standard RAM unit. Three read-write arms are provided which may be independently hunting for tracks on any of the 100 surfaces or actually engaging in reading or writing. Where a change of disk is involved, the mean access time is slightly over $\frac{1}{2}$ second; where no change of disk occurs, the access time is approximately $\frac{1}{2}$ of this.

This seems like a most useful device for storing large quantities of material but it should be remembered that in one access time you read only *one* record. This is in contrast to the photoscopic device which reads *all* in one access time.

(4) Manipulating Information

Since computers are essentially logical devices, they must have a clean-cut and unambiguous means of referring to their stored information. Referral to this pool of information for purposes of executing the sequences of commands that constitute the program will be discussed under Control.

Commercially available data processing devices differ markedly in the organization of this information in both length and in the number system employed. These will be discussed independently in the sections below.

(a) **Number System**. The first electronic computer devised employed the number system that is most applicable where bi-stable devices are used as a computing element. Such a number system is the binary system which is built on a base of two. In this system, only the digits 0 and 1 are

permissible and any quantity can be represented by a sequence of 0 and 1. For example, the number 17.5 may be expressed in binary as 10001.1. This may be better understood by realizing that the accepted method of expressing numbers always involves a base. In the decimal system when we say 17.5 we mean

$$1 \times 10^1 + 7 \times 10^0 + 5 \times 10^{-1}.$$

The number written in the binary system above is equivalent to

$$1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1}.$$

In addition to its simple representation by computer elements, the binary system has the advantage of being theoretically the most efficient means of storing information.

For practical information processing, the binary system suffers from the difficulty of poor comprehension by humans. It is only this that has influenced computer designers to move away ostensibly from the binary system. The result has been an evolutionary process by which decimal and alphabetical systems have replaced the binary. That is to say, as far as the programmer is concerned, all input and output information is treated as decimal or alphabetical information.

There are several techniques by which this is accomplished, including the automatic conversion to and from binary in the input and output stages, computing in the binary-coded-decimal mode (each decimal digit expressed as a combination of 8, 4, 2, and 1), and the use of a special register to which the alphabetical and decimal representations are sent and from which operations are executed interpretatively in binary. A system that is used often in this field is alphanumerical, often contracted to alphanumerical. This means simply the use of a wide gamut of symbols including all 10 arabic numerals, the 26 alphabetical characters and a short list of the most commonly used punctuation marks.

Where non-numerical information is to be processed, the need for alphanumerical character representation is obvious. This may be accomplished by using two-digit symbols for non-numerical characters (e.g. 21 = A, 76 = %) or by a direct representation using 6 or 7 binary digits to represent the range of characters employed.

(b) **Word Length.** Until a few years ago, computing machines always operated with fixed word length. The total storage of the computer was divided into segments of identical length. Typical lengths were 40 binary digits, 10 decimal digits, 5 alphanumerical characters, etc. This fixed length generally worked out quite well for numerical calculations, but was much less attractive for processing non-numerical information.

Recently several systems have been developed for using word lengths

whose size is adjustable. We may now process two digit numbers, a single character symbol or a 137 character bibliographic reference. A number of clever logical features make this possible. For example, in some machines, every character has an address. In reading and writing operations, the first character address is given and each succeeding character is automatically treated until a special length symbol is detected. Numerical quantity to be operated on mathematically is separated by characters containing a plus or minus sign or by non-numerical characters. General manipulation of non-numerical information within the computer is length-controlled by the size of a special register whose length may be set at will by the programmer.

Other means of obtaining variable length are available but in not nearly as sophisticated a manner. Some machines are capable of adjusting input information. The restriction usually applies that word length of less than the internal fixed length may be accepted and manipulated as fractional words. Generally no provision is made to handle words of length longer than standard.

(5) **Control** (To activate and interrelate the various components in such a manner that the desired sequence of operations takes place.)

Thus far the data processing device has been viewed as a large mass of organized information with suitable inputs and outputs. In order for commands to be carried out to process the information, a central "agency" must be responsible for administering the task. A series of commands written in the appropriate language (see below) is loaded into the computer together with the information to be processed. Superficially, such commands are indistinguishable from any other information. A control register, set to a predetermined storage location, fetches from that location the word stored there, interprets this as a command and then proceeds to execute it. When the execution has been completed, the control register is modified and then the command stored at this modified location is fetched, interpreted and executed. Many of the commands will be decision commands. In these, an interrogation is made and, depending on a specified status in the process (e.g., two numbers are equal, a number is negative, etc.), the control register is set to one of two alternative numbers. This makes it possible for the program to branch at specified points and thus one can build complicated logical structures.

In addition to branching on decisions, another powerful characteristic of these machines is their ability to modify their own instructions. Usually this is done by taking a command and, since it may be treated like a numerical quantity, adding or subtracting a stored number and then placing it back in its original location. For example a command may read "add the contents of location 1371." If we add a one to this in the units position and replace it, the command now reads "add the contents of location 1372."

Generally we may modify as often as we like. After a fixed upper limit for modification, we will often discontinue the modification and frequently restore the original command.

A newer means of effecting this is the use of index registers. In a machine making use of index registers, the command itself is never actually modified; instead a short auxiliary register keeps count of the number of modifications and is itself added to the command at the interpretation stage. The address to be fetched is the sum of the command plus the index register and is called the "effective address". For example, in the case mentioned above, the command would continue to read "add the contents of 1371," but to this command would be added successively 0, 1, 2, etc., as the index is increased. Provision is made for automatic restoring and sensing of the terminal point of the indexing process.

Data Processing Machines: Programming

The preparation of a set of logical sequential instructions for a data processing machine is termed programming. This is a skill that is being mastered by thousands and must be acquired before one can adapt data processing machines to one's own use.

Before beginning the actual writing of the command sequence, a certain amount of preliminary planning is necessary. Most important is the development of a clear picture of the logical sequence of operations and, in particular, the consequence of the various decisions that are to be made during the operation. Such a plan can best be visualized in the form of the conventional computer flow diagram. In these charts, each group of related operations appear as single blocks. Branching commands appear separately and are assigned characteristically shaped figures. An example of a portion of a flow diagram appears in Figure 29-1.

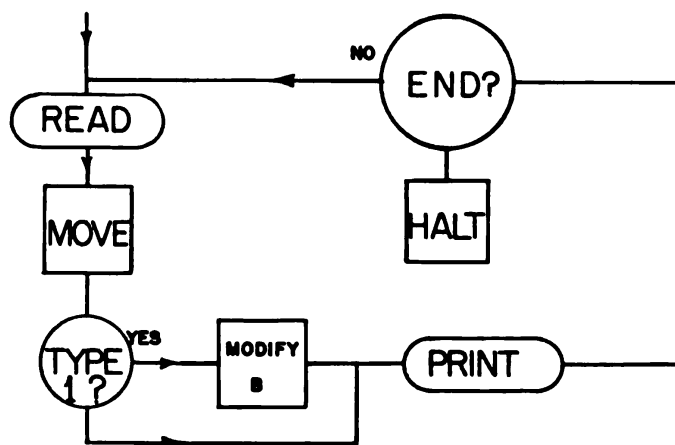


Figure 29-1. Example of computer flow chart.

After the flow diagram has been completed, the allocation of the computer storage facilities must be made. Provision should be made for the assignment of specific storage locations to the following classes of information:

- (1) The program itself
 - (a) Instructions
 - (b) Constants required in the course of the program
 - (c) Special commands used for modifying and restoring regular commands.
- (2) An input area arranged to receive raw data.
- (3) A working area in which the raw data is modified and processed.
- (4) A corresponding output area.
- (5) Tables and other information used for frequent referral.
- (6) The raw data itself, usually on magnetic tape or other slow storage.
- (7) Provision for storing the final processed output.

It is only after this layout of the flow chart and storage allocation that actual writing of the commands should begin.

In earlier times, programming was considerably more difficult and complex than it is today. The backlog of experience that has been building up has taught designer how to improve computer instruction logic and has led to the development of simplified and automatic programming. Today we classify programming methods as follows:

- (1) Direct programming in actual machine language.
- (2) Compiling of short, previously tested routines.
- (3) Symbolic assembly of programs written in a simplified language.
- (4) Interpretive programming in the course of which the machine interprets and executes pseudo-instructions of a highly simplified nature.
- (5) Many combinations of all of the above are in common use.

In *actual* programming, which should be learned first, one assembles a sequence of instructions which embodies all the logical and arithmetical steps necessary to carry out the desired processing. The advantage of programming in actual instructions lies in the speed and compactness so obtained. The disadvantage is in the repetitious nature and in the increased probability of introducing program errors.

The advantage of *compiling* lies in the avoidance of repetition of frequently used groups of commands and the simplicity with which a few instructions to the compiler may be expanded into a detailed program in machine language. Compilers also make the incorporation of miscellaneous sub-routines (e.g., input and output routines) into a larger program. Since the compiling must be done by the machine, we have added compiling time to the machine running time. In general, a program compiled by a machine will be freer of error than one written by an experienced programmer, but will be considerably less efficient in its execution time and its use of storage.

In a *symbolic* program, we assign alphabetical or numerical symbolic locations in the computer's memory (e.g., instead of storing a number in cell 1675, we write STO TALLY). The symbolic instructions are fed to the data processing machine, and the machine carries out an involved process called an assembly program, producing an output consisting of a suitable input tape (or card deck) in machine language and a cross reference listing of each symbolic command and the corresponding new command produced in machine language. This last is necessary in testing the program since, after assembly, the original symbolic addresses are of no significance. The advantages lie in the extreme simplicity of programming, the ability to insert or delete before assembly as needed and the reduced probability of making programming errors. The disadvantages of symbolic programming are the added time of assembly and the necessity for troubleshooting in machine language a program with which one is familiar only in symbolic language. Many assembly programs do provide for automatic detection for a limited number of classes of programming errors.

Interpretive programming is carried out by the use of a substitute machine language and no preliminary process of assembly or compiling is carried out. Instead, the program written in this special language is fed directly into the input of the data processor. As the calculation proceeds, each "instruction" is examined by the computer and is interpreted by referring to an internally stored set of interpretive instructions. The advantage lies in the avoidance of a separate pre-program computation and in the possibility of program testing directly in interpretive language. The disadvantages are the loss of the storage space required by the interpreting program and the slow speed induced by the necessity of analyzing each command to determine its mode of interpretation and then to execute the interpretation.

Information Retrieval Applications

While the machines described above may be used for a wide variety of processing, we are concerned here only with their use for storing and subsequently retrieving scientific and technical data. In the uses to be described, they should be regarded as natural extensions of the trend that takes us from index cards to edge-notched cards and to machine-sorted cards. All of the techniques available with these precursors may be adapted to data processing machines and, in addition, many newer and more sophisticated ones may be used. Since the applications in this area are all quite new and experience is limited, much of the material that follows will be of a suggestive or speculative nature.

In the following sections, the general principles of planning an information retrieval scheme on a data processing machine will be covered and this

will be followed by discussions of techniques for classifying and retrieving information.

General Techniques

In information retrieval schemes, it is well to distinguish between:

- (1) The actual information itself
- (2) An abstract of the information
- (3) A descriptive title
- (4) A reference to the bibliographic location of the material
- (5) A reference to the physical location
- (6) A serial number which may be cross-referenced against either 4 and 5 or both
- (7) An index according to some prearranged scheme which classifies the subject matter of the reference.

In most information retrieval systems, we have the following elements: 7, the index; 1, 2 or 3 or a combination of these; 4, 5 or 6 or a combination of these. Sometimes only the first two groups are necessary.

For illustrative purposes, we will consider a system in which each item of information contains an index and a locating reference. The necessary extension to the inclusion of further elements will become obvious with experience and familiarity with data processing systems. The data to be processed will then consist of records containing the index plus the locating records. For such records, a specific format must be devised. With some variations of this, the record length will be fixed; for others, it will be variable. For example, if the indexing scheme is a direct coded one involving the presence or absence of characteristic markings at 56 fixed locations and a 7-digit alphanumerical serial identification is used, we must arrange to have fixed sequences of 63-digit records fed to the processing machine.

After operating on the 56-digit index and determining that the associated information is to be recovered, the 7-digit locating reference is transferred to the output of the machine. Such a simple scheme corresponds exactly to the processing of punched cards through a selecting device. With the proper machine, such records may be processed at speeds of up to 10,000 per minute.

Going from this crudest operating mode to variable length records, we pass to a more sophisticated use of indexing and of the machine. In general, the amount of indexing required will vary with the nature and applicability of the primary references. Since data processing machines handle variable length records well, one might just as well capitalize on this feature when devising the index. Another way of saying this is that in previous punched card systems, we are always faced with a fixed number of holes that may be punched or notched and, no matter how simple or complex our material

is, we are committed to that number of holes. Only with variable length records on a continuous storage medium are we free to match the index length with its complexity.

There are several practical methods for handling variable length records (see 4b above). The choice of the method will, in general, depend on the particular data processing machine that is used. If the machine itself operates with variable length records, then the device used by the machine for indicating the beginning and end of such records should be used. If the machine operates with fixed length words, each record must carry with it a tally indicating the number of words that make up the record.

It is often desirable to *block* records into a larger unit. For example, 100 records may be placed together in such a block and treated as a unit. The advantage so derived involves the input to the machine. In general, it is more efficient to start and stop the input unit infrequently and to fill as much of the internal storage of the computer as possible. Blocking is quite a common practice.

Since the block itself may also be regarded as a primary record, its length will be of interest. In computing that is purely mathematical, we usually deal with blocks consisting of a fixed number of fixed-length words. In processing and retrieving information, we will often find ourselves with the opposite pole, namely a variable number of variable-length records. The techniques for forming and treating such blocks are not difficult and depend upon the use of a number of elegant techniques.

Indexing Logic for Large-Scale Retrieval Systems

In using simple needle sorting or punched card machine sorting, one generally employs a relatively unsophisticated logical searching scheme. While it is not the purpose of this chapter to provide a treatise on the logical structure of indices, some discussion is necessary to understand the scope of the use of data processing machines in this field.

Until recently, indexing was generally hierarchical. Each item to be indexed was conceived in its relationship to the whole of knowledge. During this period, such schemes as the Dewey Decimal System and the Library of Congress classification were devised. Such hierarchical schemes established fixed sequences of generic classification quite similar to the Linnean system in biology. Where it is practical to use such schemes (e.g., in a taxonomical index), they work extremely satisfactorily with a data processing machine as well as with simpler devices.

Spurred on, in part, by the advent of mechanical devices, a number of newer approaches to the indexing problem have been developed. The term "multi-dimensional" has been applied to schemes in which a referenced item is viewed as the intersection of a number of unidimensional classifica-

tion schemes. Among recent contributions to this field are the approach using symbolic logic, coordinate indexing and similar approaches to the multidimensionality of knowledge.

To retrieve information classified on these lines is a task ideally suited to the superior logic of data processing machines. A theoretical analysis of the behavior of such machines shows that all of their operations can be reduced to sequences of operations expressed in symbolic logic. Conversely, we can take a series of statements written in the notation of symbolic logic and convert them to a computer program. Granting this ability, it now becomes relatively simple to embody the logic of any of these retrieval schemes into a suitable program for a data processing machine.

Indeed, it is possible that much more complex and sophisticated schemes can be conceived for indexing since the means for practical execution is at hand. As an example, an approach to indexing based upon topological representation has been developed with the idea that data processing machines could cope with the complexity of the logic. This has indeed proven to be the case.

Hypothetical Case Study

To integrate the material that has gone before, the steps necessary to set up a data processing machine for storing and retrieving information previously present in a file of machine-sorted cards will be described in some detail.

Step 1. The operation of the particular machine to be used must be studied. Its internal capacity, its input-output facilities and the technique of programming must be mastered before any further steps can be taken.

Step 2. A re-examination of the basic information to be manipulated must be made. Since a simple one-to-one translation from cards to magnetic tape is not recommended, some fundamental questions should be reviewed. These include: the present size of the file, its rate of growth, the ultimate size, the necessity for revision of file material or classification, and the present and future rate of utilization of the file. With the increasingly rapid distribution of technical information, one is well advised to allow for growth in all directions in setting up a system.

Step 3. One must decide upon suitable formats for indices, indexed material and questions posed for retrieval. The format may be either fixed or variable, depending upon the type of material and of indexing that is used. If it is fixed, the exact number of fields to be used and the number of characters to be allotted to each field must be settled. If variable, the means of signaling to the computer the length of the variable fields must be decided upon and the variable field format planned in detail.

Step 4. After settling the format question, the logic to be used in in-

dexing and retrieval must be re-examined to determine what sequence of computer operations can be used to file and retrieve information according to this scheme.

Step 5. With this logical plan in mind, the flow chart for the entire retrieval scheme should be laid out. In carrying this out, many logical inconsistencies will come to the surface and considerations previously missed will be discovered. In planning the flow chart, all of the possible abnormal operations detected by the data processor should be considered and provisions made for decisions when they occur. Some of these conditions are—input and output errors, impossible commands to the machine, inconsistent commands, referral to non-existent or improper memory locations, etc. Provision should also be made to report should any conditions not foreseen in the retrieval process occur. These will include missing items, incomplete items and conditions inconsistent with the logic of the indexing method.

Step 6. The preparation of the program itself should now take place. The use of symbolic programming with machine assembly is recommended where possible. After completing the first draft of the program, the beginner is urged to check it over repeatedly, since the operating cost of most data processors is ten to one hundred times the hourly labor cost. One can profitably afford to spend many hours checking work that will run for less than one hour on the machine.

Step 7. The original file of data on punched cards can be converted directly to magnetic tape without editing, with one exception. Most card-to-tape converters will signal an error if certain forbidden combinations of punches occur. It is generally recommended that the punched cards be run through a collator or sorter to detect and correct these errors before taking the cards to the data processing machine room.

Step 8. After the data cards have been directly transcribed to "raw" tape, the original punched card format (now on tape) must be converted to the format selected for the new operation. This requires a separate editing program for this conversion which, in general, is much simpler than the retrieval program. One should be careful to distinguish the program for converting the raw tape to the actual retrieval tape from the program for the retrieval itself.

Step 9. When the program has been checked manually to complete satisfaction, it should be punched onto cards for loading into the machine. In punching, care should be taken that the trouble spent in checking the program is not wasted by careless key punching. All key punching should be verified and then listed for ready reference. When using combined numerical and alphabetical information, a common pitfall is the confusion between zero and the letter O and between one and the letter I.

Step 10. After the cards are punched, if they are to be assembled or compiled, they must be fed to the computer for this process. The output of the machine will generally be a deck of cards in actual machine language and a cross-reference table relating the symbolic commands with the actual ones.

Step 11. Now, with the edited retrieval tape on a tape reader, the program cards are placed in the card reader and the first attempt at running it is made. The novice can expect complete chaos. I quote the following from *Computing News*, Vol. 3, August 1, 1956.

"The person who goes on to a machine for his first debugging session should be prepared for a shock. Particularly if he is conscientious in his programming and has carefully checked the keypunching, and has sorted and sequence checked his deck, and has had no troubles in assembly—his problem is NOT going to run smoothly and he may as well face it. It seems to be a sad truth that no one can write fifty lines of instructions or so without making some sort of error; these errors the machine will uncover almost immediately. His claim at this point "But I checked it!" has no apparent effect on the machine. Fortunately, nearly every possible programming error results in fairly clear-cut indication on the console lights when the machine stops.

"After having experienced several times the ability of the machine to locate your errors unerringly (so to speak), the first occasion when a problem runs clear through seems equally shocking. Usually the results make little sense at this stage, and the debugging process continues.

"The first few go-arounds at the machine rather resemble an initiation: the victim is given a shove down a line of knowing devils, all equipped to heckle and harass. He emerges at the other end shaken, confused, and slightly hurt. Similarly with the machine. Various persons are on hand, all with vast experience (i.e., more than you've had). Orders are shouted in rapid-fire:

'Load the reader.'
'Push some buttons.'
'Run out your cards.'
'Eject your listing.'
'Get your punched deck.'
'Get off the machine.'
'Get out.' "

Step 12. After the initial trial, which will be unsuccessful in 99.9 per cent of the cases, the process of "debugging" begins. Experience shows that this will require a number of short runs or attempts on the machine, interspersed with considerably longer periods of confusion followed by enlight-

enment concerning the latest difficulties encountered. When the trouble has been found, one is often at a loss to determine the most efficient way of inserting the needed correction. One of the most common sources of frustration is the frequency of inserting mistakes in the process of correcting another error.

Step 13. Eventually, the program will reach the point where it runs all the way through. This is by no means the end of the debugging. By reference to the flow chart, one finds that only a few of the alternate logical paths have been traversed. While it is frequently a challenge to introduce one of the requisite conditions, no program should be regarded as debugged until every path has been checked out. Even when all of the paths have been traversed, one cannot be sure that a variable length program will operate correctly until records of the smallest and largest size have been processed. Indeed, it is quite a challenge to devise a series of tests that will completely check out the program.

Step 14. At this point, the retrieval process may begin and can be carried out by the machine. During the first runs, considerable care should be taken to see that the actual results obtained are those desired. If not, the program may be operating perfectly, but the original logic or its adaptation to the machine may be faulty and require revision.

Step 15. After the system has been operating satisfactorily, one will soon encounter new references to be added to the information tape, as well as corrections and deletions. Now it will be necessary to write a new program called an "up-dating" program. This usually provides for any changes in the tape to be fed in on punched cards. As the cards feed in, the original tape is simultaneously (and synchronously with respect to the information sequence) fed to the computer. As this process goes on, a new tape is written which merely copies portions of the old tape that are to be written intact. Any material that is to be changed is modified within the data processor and written on the tape in its new form. Like the retrieval program and the raw tape editing program, this, too, must be carried through the programming, card preparation and debugging and testing stages.

Step 16. From time to time, the up-dating program should be run to maintain the information file in as perfect a form as possible.

General Bibliography

A. Review Article

Heumann, K. F., "Data Processing for Scientists," *Science*, 124, No. 3226 (October 26, 1956).

B. On Data Processors

Anthony, R. N., Ed., "Proceedings Automatic Data Processing Conference," Boston, Harvard University Press, 1956.

Canning, R. G., "Electronic Data Processing for Business and Industry," New York, John Wiley & Sons, Inc., 1956.

Dooher, M. J., Ed., "Electronic Data Processing in Industry," New York, American Management Association, 1955

Haskins and Sells, "Data Processing by Electronics," New York, Haskins and Sells, 1954.

Klingman, H. F., Ed., "Electronics in Business," New York, Controllershship Foundation, 1955.

Marting, E., Ed., "A New Approach to Office Mechanization: Integrated Data Processing through Common Language Machines", New York, American Management Association, 1954.

C. *On the Logic of Retrieval Methods*

Moores, C. N., "Information Retrieval on Structured Content," Third London Symposium on Information Theory, Sept. 12-16, 1955.

Opler A., and Norton, T. R., *Chem. Eng. News*, 34, 2812 (1956).

Opler A., and Norton, T. R., "A Manual for Programming Computers for Use with a Mechanized System for Searching Organic Compounds," Dow Chemical Co., Pittsburg, Calif., 1956.

Perry, J. W., Kent, A., Berry, M. M., "Machine Literature Searching," (Western Reserve Press; Interscience, New York, 1956), annotated bibliography.

Taube M., *et al.*, "Studies in Coordinate Indexing," Washington, Documentation, Inc., 1953-: vols. I-IV have appeared.

Part V

ANNOTATED BIBLIOGRAPHY ON USES OF PUNCHED CARDS

Chapter 30

ANNOTATED BIBLIOGRAPHY ON USES OF PUNCHED CARDS

ROBERT S. CASEY

W. A. Sheaffer Pen Company, Fort Madison, Iowa

The numbering follows the 276 item bibliography in the first edition.

The following initials at the end of an entry indicate the source of the reference, which usually carries an abstract of the article:

A.D.-:-. *American Documentation*, Volume: Number

C.A.-:-. *Chemical Abstracts*, Volume: Column

C.L.-:-. *Chemical Literature*, Volume: Number

L.K. Loftus-Kent bibliography, item No. 500 below

CLASSIFICATION

Method; Substance; Treatment

- A** Cards, Plain
- B** Cards, Edge-marked
- C** Cards, Edge-notched (not pre-punched)
- D** Cards, Edge-punched (pre-punched holes)
- E** Cards, Center-punched
- F** Coding
- G** Coordinate indexing
- H** Experimental; Research; Speculation; Recommendation; Proposal
- J** Machines, Punched card
- K** Machines, Special purpose
- L** Microfilm; Photography
- M** Nomenclature; Notation; Terminology
- N** Review; General; Broad; Critical
- O** Subject Analysis; Concepts; Organization of Information
- P** Systems
- Q** Theoretical

Application

- 1** Bibliography; Documentation in General
- 2** Calculation; Correlation; Analysis of Data
- 3** Chemistry, Analytical
- 4** Chemistry, Inorganic
- 5** Chemistry, Nuclear
- 6** Chemistry, Organic
- 7** Engineering
- 8** Industry; Economics
- 9** Information Storage and Retrieval
- 10** Law
- 11** Library
- 12** Medicine; Biological Sciences
- 13** Patents
- 14** Physics; Physical Phenomena
- 15** Reports
- 16** Science; -other; or -in general
- 17** Translation; Linguistics

277. A.B.A. Operations Research Committee, "Sub-committee on Punched Cards and Legal Literature Searching," Mimeographed monograph no. 1—Nature and Scope of Program, 28 June 1954. Subjects for study will include symbolic logic, the theory of games, the information or communication theory, electronic computers and automatic literature searching devices. L.K. NQ9 10
278. "The Accessibility of Knowledge II," *Endeavour*, 11 (41), 3,4 (1952). This paper considers some means which have been devised to assist the research worker in his search for information. C.L.4:2. N1
279. Aikele, Erwin, "Importance of Electronic Computers and their Possible Uses in Documentation," *Nachr. Dokument.*, 3, 26-31 (March 1952). C.L.4:3. KN1
280. Aikele, E., H. Christen, *et al.*, "Handbuch der Lockkarten-Organization," Agenor Druckund Verlags-GmbH, Frankfurt am Main, 240 pp. (1957). The principles of punched cards and their installations are described. One hundred illustrations. C.L.9:3. N
281. Alexander, S. N. (Data Processing Systems Division, National Bureau of Standards, Washington, D. C.), "Data-Processors for Information-Retrieval Purposes." Presented at the 132nd national meeting of the American Chemical Society, New York City, September 1957. New areas of investigation with data processing systems are reviewed and the implied trend in equipment requirements is discussed. C.L.9:3. JKN9
282. Amble, E., Per Andersen, and H. Viervoll, "Punched-card Method for Radial-Distribution Calculations in X-Ray and Electron-Diffraction Studies", *Acta Chem. Scand.*, 5, 931-5 (1951) (in English). C.A.46:9973. J2 14
283. Andrews, D. D., "Interrelated Logic Accumulating Scanner (ILAS)," U. S. Patent Office Research and Development Rept. No. 6, (June 1957). A description of the punched card multicolumn sorter designed for use by the Patent Office. C.L.9:4. K9 13
284. Andrews, D. D., (Research and Development, United States Patent Office, Washington, D. C.), "Major Factors Governing the Selection of Machine Data-Handling Technique". Presented at the 132nd national meeting of the American Chemical Society, New York City, September, 1957. The experiences of the Research and Development Division of the United States Patent Office in selecting machine techniques of literature control are briefly reviewed. C.L.9:3. HJ13
285. Andrews, D. D., and S. M. Newman, "Storage and Retrieval of Contents of Technical Literature, Non-chemical Information," Office of Research and Development, U. S. Patent Office, Dept. of Commerce, Washington, 12 pp. (1956). Discusses coding of concepts. O9
286. Andrews, E. R., F. Fallon, G. L. Krueger, F. P. Palopoli, E. L. Schumann, and K. W. Wheeler, (Dept. of Organic Chemistry, The William S. Merrell Company, Cincinnati, Ohio), "Structure Code for Organic Compounds." Presented at the 131st national meeting of the American Chemical Society, Miami, Florida, April, 1957. The molecular formula, number and kind of rings, length and type of unsaturation of alkylene and terminal chains, functional groups, and special molecular properties (such as chelates, solvates, organometallic compounds) are described in 36 columns on IBM punched cards. An additional eight columns are assigned to indicate the source, serial number, lot number, and specific salt of the compound. C.L.9:1. FJM6
287. Ashe, Geoffrey, "Introducing Computers to Beginners," *Computers and Automation*, 3, 8-11 (March 1954). Stresses the need for more purely orientation material in introductory courses on computers. A.D.5:4. KN
288. Ashthorpe, H. D., "The (Machine) Punched Card Indexing Experiment at the

- Library of the Atomic Energy Research Establishment, Harwell," *ASLIB Proceedings*, 4, 101-104 (May 1952). Discusses the experiment and reasons for reverting to a visual index. C.L.5:3 L.K. **AHJ**
289. "ASTIA's New System of Indexing," *Library J.*, 78, 2080-2081 (1953). A completely automatic indexing machine, which will present document numbers in answer to reference questions typed on a conventional typewriter keyboard, is under development for use with the Uniterm System. C.L.6:2. **GK**
290. Astin, A. V., "Computing Machines and Automation," *Computers and Automation*, 5, 6-11, 38 (April 1956). In his testimony before the Subcommittee on Economic Stabilization, October 27, 1955, Dr. Astin discusses briefly automation and various phases of mechanization, the most recent of which is the utilization of advanced and automatic computational techniques principally in the field of science and technology. He describes the work the Bureau is doing in collaboration with other government agencies to develop new uses for the modern computing machines. A.D.7:4. **KN16**
291. "An Automatic Microimage File Searches Film Record and Makes Photographic Print," National Bureau of Standards, *Technical News Bulletin*, 40, 89-90 (July 1956). A data storage and retrieval device developed by M. L. Kuder of the Electronic Instrumentation Laboratory, Bureau of Standards, provides rapid access to any one of 10,000 information-containing frames recorded in miniature on a 10-inch square sheet of microfilm. A.D.8:1. **KL9**
292. "An Automatic Micro-Image File," *Computers and Automation*, 5, No. 11, 26-27, 50 (November 1956). A description of a micro-image data storage and retrieval device developed at the National Bureau of Standards to provide rapid access to any one of 10,000 information-containing frames recorded in miniature on a 10-inch square sheet of microfilm. A.D.8:2. **KL9**
293. Avakian, Emil A., "Initial Report on the Automatic Micro-Film Information System (AMFIS)," Mimeographed report (2 September 1952). See following reference. A.D. **KL9**
294. Avakian, Emil, A., and Eugene Garfield, "AMFIS—The Automatic Microfilm Information System," *Spec. Libr.*, 48, 145-148 (April 1957). An electronic microfilm selector and projector is described which can retrieve and project any one of several million microfilm frames by dialing the call number. C.L.9:3. **KL9**
295. Baatz, Wilmer H., and E. H. Maurer, "Machines at Work," *Library J.*, 78, 1277-81 (1953). Description of the adaptation of punched card machinery to special projects, as well as library routines, at Milwaukee Public Library. C.L.5:4. **J11**
296. Bailey, M. F., B. E. Lanham and J. Leibowitz, "Problems of Classification and Documentation in the U. S. Patent Office in the Field of Petroleum and Allied Subjects," Proceedings Third World Petroleum Congress, Section X, 14-21 (1951). Describes machine searching (I.B.M.) of patents for compositions of matter in terms of ingredients or functions or both. Discusses principles developed in formulating the Patent Office classification. **JO6 13**
297. Bailey, M. F., B. E. Lanham, and J. Leibowitz, "Mechanized Searching in the U. S. Patent Office." (Presented at a meeting of the Division of Chemical Literature, American Chemical Society, February, 1951.) *J. Patent Office Soc.*, 35, 566-587 (August 1953). Describes an experiment in mechanized searching undertaken by the Patent Office in the field of medicinal compositions. Punched card codes were provided for a number of simultaneously used systems of subject categories. L.K. **FHJ12**

298. Bar-Hillel, Yehoshua, "Can Translation be Mechanized", *Am. Scientist*, **42**, 248-60 (1954). A discussion of the difficulties involved with special application to German-to-English translation. C.L.6:3. 17
299. Bar-Hillel, Yehoshua, "The Present State of Research on Mechanical Translation," *Am. Doc.*, **2**, 229-237 (October 1951). Discusses the objectives of machine translation, divisions of labor between man and machine in a translation partnership, and techniques which must be developed before more effective mechanical translations can become possible. A.D. 17
300. Bar-Hillel, Yehoshua, "Machine Translation," *Computers and Automation*, **2**, 1-6 (July 1953). Discussion of the present state of research on machine translation at the Research Laboratory of Electronics, supported by the Signal Corps, the Air Material Command, ONR, and the Rockefeller Foundation. L.K. 17
301. Batchelder, R. W., "The Microcard," *Spec. Libr.*, **43**, 157-61 (1952). A review of developments in microfilm and microcards including microcard readers. C.L.4:3. LN
302. Begun, George M., "Making Your Own Punched Cards," *J. Chem. Educ.*, **32**, 328 (1955). Uses template made from galvanized iron 5" x 8", corner beveled, holes slightly larger than those in finished card. Four or five file cards are lined up with template, corners cut with scissors, holes cut with hand punch. D
303. Bello, Francis, "The Information Theory," *Fortune*, **48**, No. 6, 136-41, 149-50, 152, 154, 156, 158, (1953). Q
304. Benson, Frederic R., (Remington Rand, Inc., South Norwalk, Conn.) "Recording and Recovering Chemical Information with Standard Tabulating Equipment." Presented at the 124th national meeting of the American Chemical Society, Chicago, Illinois, September 1953. C.L.5:3. J6 9
305. Benson, B. S., "Description of an Automatic Data-Analyzing Machine," *Mech. Eng.*, **72**, 285-88 (1950). This paper describes a machine developed by the Douglas Aircraft Company for the automatic analysis of bar-graph records. C.L.3:4. K2
306. Bergmann, G., and G. Kresze, "Card-Index for Documentation in Molecular Spectroscopy," *Angew. Chem.*, **67**, 685-94 (1955). Edge punched cards are used for documentation of literature on molecular spectroscopy. Organic substances are coded on the basis of their composition. C.L.8:1. D1 6 14.
307. Berkeley, Edmund C., "The Relations between Symbolic Logic and Large Scale Calculating Machines," *Science*, **112**, 395-9 (1950). KOQ
308. Berry, Madeline M., and P. Jenkins, "A Punched Card Index to a File of Documentation References." Presented at the 125th national meeting of the American Chemical Society, Kansas City, September 1954. A file of preprints and reprints on the collection, storage and retrieval of scientific literature, as well as general bibliographies, manufacturers' bulletins, brochures, etc., is indexed by means of a simple hand-sorted punched card system. C.L.6:1. D1 9
309. "Bibliography on Electron Microscopy," *Science*, **118**, 378 (1953). Announcement that the New York Society of Electron Microscopists is to issue a bibliography on edge punched cards of the world literature in all fields of electron microscopy. D1 14
310. "Bibliography on Filing, Classification and Indexing Systems for Engineering Offices and Libraries," *ESL Bibliography*, No. 9. Engineering Societies Library, New York, 18 pp (1954). Includes lists of subject headings, and hand-sorted, punch-card systems suitable for organizing small collections of items. C.L.6:4. DOI 7

311. "Bibliography on the Use of IBM Machines in Science, Statistics, and Education," International Business Machines Corp., New York, 37 pp. (1950). J1 16
312. Bien, Saul M., "Registration of Orthodontic Diagnostic Records for Statistical Evaluation," *Am. J. Orthodontics*, 41, 482-3 (1955). An abstract of the paper presented before the Research Section, Am. Assoc. Orthodontists, Chicago, Illinois, May 9, 1954, and illustration of the 8" $\times$ 10½" edge-punched card. D2 12
313. Biunno, Vincent P., "Searching Legal Literature—An Appraisal of New Methods," *Law Library J.*, 46, No. 2, 110-19 (1953). Includes discussion of punched cards. DJN10
314. Biunno, Vincent P., "Informal Papers on Legal Literature Searching," New Jersey Law Institute. Includes discussion of automatic searching devices. L.K. DJN10
315. Bjorkbom, C., "New Views on Indexing: Post-Combination Systems," *Tidsskr. Dok*, 11, No. 4, 41-44 (1955). Several such systems are described: Batten system, Uniterm, Cordonnier, Selector, Peek-a-boo cards and the system used at the National Bureau of Standards. C.L.8:3. EN
316. Blackman, Nelson, M., "-ACS", *Am. Scientist*, 44, 198-203 (1956). A survey of the names and etymologies of automatic digital computers ("-ACS"). C.L.8:4. K
317. Blackman, Nelson, M., "Eni-, Meni-, Mini-, Moni-ACS," *Instrumentation and Automation*, 30, 1302-03 (1957). The names and abbreviations of many well known and some less familiar (incl. British and French) automatic digital computers are explained. C.L.9:4. K
318. Bloetjes, Mary K., "Management Research in Food Service Operations." Presented at the 37th Annual meeting of the American Dietetic Association, Philadelphia, October 29, 1954. Suggests studying the adaptation of edge punched cards to quantity food service operations. D8
319. Bohnert, L. M., "Two Methods of Organizing Technical Information for Search," *Am. Doc.*, 6, 134-151 (July, 1955). One is the traditional methods of library classification; the other non-heirarchical, relying on combinations of general terms to characterize specific ideas or terms. The activities of seven groups are discussed: (1) the NRC Chemical-Biological Coordination Center, (2) the U. S. Patent Office, (3) the Welch Medical Library, (4) J. W. Perry and associates, (5) the NBS Office of Basic Instrumentation, (6) the Zator Co., and (7) M. Taube and associates. L.K. NO
320. Bolton, Houcke E., "A System for Cardiac Studies," *Systemes*, 19, 19 (March-April 1955). Describes the use of punched cards to facilitate statistical studies of case histories involving cardiac surgery. A.D.6:3. J2 12
321. Bonino, J. J. and K. M. Laing, "Punched Card Classification of the Nuclides," *Nucleonics*, 12, No. 2, 68 (1953). Elements' symbols, atomic number, mass number, half life, energies of various emissions, stability, kind of emission, availability from AEC, and presence of natural radioactivity, are coded on edge punched cards. C.L.5:4. D5
322. Booth, Andrew D., "Mechanical Translations," *ASLIB Proc.*, 9, 177-81 (June 1957). C.L.9:4. 17
323. Booth, Andrew D., and K. H. V. Booth, "Automatic Digital Calculators," Academic Press, Inc., New York, 1953. A guide to the theory of digital calculators, meant particularly to introduce new workers in the field to the notions, language and literature of the subject. A.D.5:4. KNQ
324. Boundy, R. H. (Dow Chemical Co., Midland Mich.) "A Research Director's Views on Literature Control. What are the Possibilities of Application of

- Data-Handling Equipment?," Presented at the 132nd national meeting of the American Chemical Society, New York, September, 1957. C.L.9:3. N
325. Braband, C., "Some Proposals for the Development of Visually Inspected Punched Cards," *Nachr. Dokument*, 8, 42-44 (March 1957). System of super-imposable, visually inspected punched cards associated with the names of Batten, Cordonnier and others. C.L.9:4. EH
326. Brillouin, Leon, "Science and Information Theory," Academic Press, Inc., New York, 320 pp, 1956. Definition of information; principles, problems, and applications of coding; analysis of signals; thermodynamics; negentropy principle; Maxwell's demon; writing, printing, and reading; computing. C.L.8:4. FNOQ
327. Bristol, Roger Pattrell, "Can Analysis of Information be Mechanized?," *College and Research Libraries*, 13, 131-135 (April 1952). Reviews the problem of making subject analysis amenable to mechanization and the development of machines capable of recording, analyzing and retrieving this information. L.K. NO
328. Bristol, Roger Pattrell, "Control of Subject Information: Can it be Mechanized?," *College and Research Libraries*, 11, No. 3, Part I, 222-27 (1950). Review of tools being used for the recording and releasing of subject information. C.L.2:4. N9
329. Brown, Russell V., "Developing and Using Numbering Systems", *Machine Design*, 28, 106-8, (17 May 1956). Topics discussed are: developing coding systems; effect of coding; straight sequence code; block codes; group classification; significant digit codes; final and decimal codes; mnemonic symbols; numeric-alphabetic codes. A.D.7:4. F
330. Brown, Curtis L., (National Research Council, Washington, D. C.) "Two Approaches to the Retrieval from a Special Library. Part I. Application of a Simple Punched Card System to a Special Information Service." Mortimer Taube (Documentation, Inc., Washington, D. C.) "Part II. Application of the 'Uniterm' System and the 'Association of Ideas' to a Special Library." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. The Prevention of Deterioration Center of the National Research Council has installed a direct-code Keysort punched card system, which is both simple and economical enough to be of interest to small technical groups and individuals. The operation, advantages, and limitations of this system are presented in the first part. In the second part, the potentialities of the "Uniterm" indexing system and of the mechanized "association of ideas" are discussed as they could be applied to the same library material. C.L.7:3. DG11
331. Brown, R. Hunt, "Office Automation," Automation Consultants, New York, 283 pp., xviii (1955). Loose-leaf \$25.00 per year service to keep manual up-to-date. Non-technical description of integrated and electronic data processing equipment and methods. C.L.8:1. JK
332. Brown, George W., Gilbert W. King, and Louis N. Ridenour, "Considerations Bearing on the Use of Photographic Techniques for Information Storage," Mimeographed report (26 March 1953). The proper exploitation of photographic techniques should yield useful storage systems having total capacities in the range from 10 to 100 megabits of information. With proper indexing, it should be possible to gain random access to any entry in a time comparable to that which is needed to look up a random address on the commonly used magnetic drum. Flying spot scanner is discussed, in combination with magnetic as well as photographic storage. L.K. HKL

333. Brown, W. F., and G. Oneal, "Library Searches with Punched-Card Machines," *Science*, 123, 722-23 (1956). The abstracting, coding and searching procedures with machine methods as employed at Sun Oil Company's Physical Laboratory are explained. C.L.8:3. J9
334. Burkhard, C. A. (Research Laboratory, The Knolls, General Electric Co., Schenectady, N. Y.) "Storing and Finding Technical Information with Punched Cards." Presented at the 124th national meeting of the American Chemical Society, Chicago, Ill., September 1953. Review of hand- and machine-sorted punched cards for bibliographical, data assembly, correlation, and analytical problems. C.L.5:3. DJ1 2
335. Busa, R., "Mechanization Applied to Linguistic Analysis," *Nachr. Dokument.*, 8, 20-26 (March 1957). Details of IBM machinery applied in the linguistic analysis of ancient texts. C.L.9:4. J17
336. Bush, V., Chairman, "Report to the Secretary of Commerce by the Advisory Committee on Application of Machines to Patent Office Operations," Department of Commerce, Washington, D. C., 76 pp. (1954). The Committee's specific recommendations are: (1) Patent Office should put machine searching on an operational basis, (2) reclassification of patents should be accelerated, (3) establish a research and development unit in the Patent Office, (4) undertake a joint program with the National Bureau of Standards to stimulate and develop machines and techniques adapted to Patent Office operations, (5) set up an Advisory Committee in the Office of the Secretary of Commerce to stimulate and coordinate the program and related efforts within the Commerce Department. C.L.7:3. HN13
337. Campbell, D. J., "The Use of Notches in Cards as a Means of Signalling Information," *J. Document.*, 9, 224-225 (1953). Suggestion for use of various sizes and shapes of notches in the upper edges of catalog cards for signalling information about the works covered. C.L.6:2. C
338. Canning, R. G., "Electronic Data Processing for Business and Industry", John Wiley & Sons, Inc., New York, 332 pp., 1956. Discusses application of digital data processing devices to business clerical functions and answers such questions as what electronic data processing is, what it can do for a company, and how a company should go about investigating it for use in its concern. C.L.9:2. A.D.7:4. K8
339. Carwardine, J. E. D., "A Punch Card System," *Chemistry in Can.*, 2, No. 9, 36-38 (1950). An indexing system is described for an organo-chemical type of industry from the process chemist's point of view. C.L.2:4. D6
340. "Case Histories on Cards," *Chem. Week.*, 76, No. 5, 58, 60, 62 (1955). Edge punched cards are used by Standard Steel Corp. for coding, retrieving and correlating engineering and cost data. C.L.7:3. D2 7
341. "Check Handling System Employs Magnetic Ink," *Am. Ink Maker*, 34, No. 11, 57 (1956). Magnetic ink character recognition machine, developed by Stanford Research Institute is recommended as most suitable for mechanical check handling by Technical Subcommittee on Mechanization of Check Handling of Am. Bankers Assoc. Machine "reads" Arabic numbers printed in magnetic ink containing particles of iron oxide. K8
342. "Circulation Records Control," *The Punched Card Annual*, 2, 89-93 (1953-54). Describes use of IBM cards as used at Montclair Public Library, Montclair, New Jersey. C.L.7:1. J11
343. Clarke, S. H., "A Use for the Paramount Sorting System", *Nature*, 137, 535 (1936). Edge punched cards serve as a three-dimensional graph when numerical values, direct coded along edges, are sorted into order. D2

344. Clason, W. E., "The Classification of Literature by the Uniterm System," *Nachr. Dokument.*, 8, 27-29 (March 1957). C.L.9:4. **G1**
345. "Classifying the Carbohydrates," *Chem. Eng. News*, 35, 90 (June 24, 1957). A simple classification system for the carbohydrates has been developed by H. S. Isbell at the National Bureau of Standards. C.L.9:4. **O6**
346. Cline, Carl A. (Research and Development Dept., Continental Oil Company, Ponca City, Oklahoma) "An Edge-Punched Index Card System for Mechanical Sorting." Presented at the 126th national meeting of the American Chemical Society, New York, September 1954. Subjects are coded in four alphabet scales around the periphery of a 4" x 6" edge punched card as superimposed four-letter code words. The code words are derived from the subject words by removing the letters which have the highest frequency of occurrence in English words. C.L.6:3. **DF**
347. "Coding for Engineering Speed," *Chem. Eng. News*, 33, 1720 (1955). The Midwest Research Institute has established a code library containing computers' instructions for problems in chemical, civil, electrical, and mechanical engineering. With codes on file, a client for engineering computations may need to pay only for the actual time spent for the operations done on the calculator plus a nominal per cent which is credited to the development of the code library. A.D.6:3. **FK7**
348. Coile, R. (Operations Evaluation Group, Massachusetts Institute of Technology Alexandria, Va.) "Application of Concept Coordination to Document Control," Presented at the 132nd national meeting of the American Chemical Society, New York, September, 1957. Recommends an operations research before undertaking mechanized data handling. The optimum mechanization system should be useful for document control as well as for information retrieval. An application of concept coordination using computer facilities is suggested for research literature collections. C.L.9:3. **GHKO**
349. "Computers," *Chem. Eng.*, 57, 117-30 (1950). A general review of information on high-speed computers, including also some data on edge-notched cards. C.L.3:1. **DKN**
350. "Computing News," (Mimeographed. About 3 issues per month). Numerical Analysis Laboratory, Univ. of Wisconsin, B-9 Bascom Hall, Madison 6, Wis. (1953). No charge. Send own stamped, self-addressed envelopes. Punched card and calculating machines. News. Brief technical items, such as plug-board wiring. **JK**
351. "Computer Directory 1955," *Computers and Automation*, the entire issue (June 1955) A three part directory: "Part 1: Who is Who in the Computer Field" contains about 7500 entries; "Part 2: Roster of Organizations in the Computer Field" contains 300 entries; "Part 3: The Computer Field: Products and Services for Sale" contains about 600 entries classified under such headings as Adding Machines, Analog Computers, Card-to-Tape Converters, Computer Services, Consulting Services, Data Processing Machines. A.D.6:4. **KN**
352. Conroy, W. Allen, *et al.*, "The Chicago Keysort Anesthesia Record", *Anesthesiology*, 9, 121-33 (1948). Edge punched cards. **D12**
353. "Co-ordination Abstracts to be Pooled," *Chem. Eng. News*, 31, 1102 (1953). News item reporting that Alan L. McClelland, Chem. Dept., Univ. of Conn., is soliciting cooperation in a plan for pooling hand sorted punched cards bearing abstracts on metallic ion coordination compounds. **D4**
354. Corby, Roy A., Harold J. Behm, and Alvin T. Maierson, Compilers and Editors. "A Machine System for Accepting, Storing, and Searching Engineering Data on Electronic Components," Wright-Patterson Air Force Base, Ohio; Wright

- Air Development Center (March 1954). The report describes the elements of a machine-sorted punched-card system for recording, searching, and tabulating data on any electronic component. A.D.6:4. J7 9
355. "Correspondence Regarding Metallurgical Documentation and the Cordonnier-Batten System of Punched Cards", *Monthly Circular for Members of Unesco committees and collaborating bodies concerned with Documentation of the Natural Sciences*, pp. 6-11, 15 (November 1952). A.D.4:2. E7
356. "Corrosion Literature," *Chem. Inds. Week*, 68, No. 16, 25 (1951). Edge punched cards will be used by the National Association of Corrosion Engineers to speed the study of literature on corrosion data. C.L.3:3. D7
357. Cox, E. G., "Punched-card Methods in Crystal Structure Analysis," *Computing Methods and the Phase Problem in X-Ray Cryst., Anal., X-Ray Cryst. Anal. Lab., Dept. Phys., Penna. State Coll.* 1952, 132-40; cf. C.A.42: 7627g; 46:5959c. C.A.48:4963. J2 14
358. Creitz, E. Carroll, "National Research Council—National Bureau of Standards Infrared Punch Card System," Mimeographed circular, 4 pp. plus Code, 5 pp. National Bureau of Standards, Washington (1952). A punched card system for the exchange of infra red spectral data between labs. on a cooperative basis. The system uses $6\frac{1}{2}$ " x $7\frac{1}{2}$ " edge punched cards. On the compound card is coded (1) position of major absorption bands (2) melting or boiling point, (3) molecular functional groups, (4) number of C atoms for the compound in question. Serial number, technical data and absorption vs. wave length curve are shown on the card. The bibliography card shows a reference and abstract. Coded are author, reference number (from a 10,000 name numerical index) subject classification, year of publication, and type of instrument used. D1 14
359. Cullmann, Ralph E. (Letter to the editor) *J. Chem. Educ.*, 30, 246 (1953), Describes making ones own punched cards by using a punch of the type used in preparing papers for plastic binders. D
360. Cummings, Carl E. and Jack Sherman, "Statistical Analysis of Experimental Data by Means of Punch Cards," *Chem. Eng. Prog. Symposium Series No. 8*, Van Antwerpen, editor, 49, 43, American Institute Chemical Engineers, New York (1953) C.A.48:2947. J2 7
361. Daily, Jay E., "A Notation for Subject Retrieval Files", *Am. Doc.*, 7, No. 3, 210-14 (1956). Uses letters, numerals and symbols available on standard typewriter keyboard. FM9
362. "Data Processing Digest," Vol. 1, No. 1, January 1955, published monthly by Canning, Sisson and Associates, 914 So. Robertson Blvd., Los Angeles 35, California. Contains short digests of articles on office automation, electronic computation, and operations research with descriptions of programming and equipment for data processing systems. A.D.8:2. JK
363. David, E. E., Jr., "Voice-Actuated Machines: Problems and Possibilities," *Bell Labs. Record.*, 35, 281-86 (1957). C.L.9:4. K
364. Davis, L. R., "Punch Card File System," (for slides and negatives), *U. S. Camera*, 16, No. 9, 68-9 (1953). Suggests making a list of general subject words, and direct coding the subject of each picture on edge punched cards. Each card bears the serial number of one of the negatives which are stored in serial order. D9
365. Davis, B., "Control and Storage of a Slide-File Collection," M. S. Thesis, Simmons College, Boston, Mass. (1956) C.L.9:2. 9

366. Davis, Albert S., Jr., "The Legal Aspects of Machine Documentation," *Spec. Libr.*, 44, No. 1, 20-22 (1953). Points out the pitfalls inherent in delegating an essentially human function, judgment, to a machine. C.L.5:9. **JKN**
367. DeCarlo, Charles R., "The Future of Automatic Information Handling in Chemical Engineering," *Chem. Eng. Progr.*, 51, 487-91 (1955). Fundamentals of equipment now being developed for application to research, design, economic, production, and automation problems. Photographs. C.L.8:4. **JK7**
368. Dewar, D., "Preparation of Linear-Function Tables on a Hollerith Tabulating Machine," *Meteorol. Mag.*, 79, 137-40 (1950). Applied to upper air mean temperature tables. **J2 14**
369. Dewar, D., "The Hollerith Card System Applied to Upper Air Data," *Meteorol. Mag.*, 78, 163-6 (1950). **J2 14**
370. Dinwiddie, S. W., and C. C. Conrad, "Report Indexing by Hand-Sorted Punched Cards" Chapter 16 in "The Technical Report" Weil, ed. Reinhold Publishing Corp., N. Y. (1954). Superimposed coding of subjects in four rows on 5" x 8" hand sorted punched cards. The top and bottom edges of the card are sorted separately. A special four-letter code was prepared based on a 29-character alphabet. A caption-code key is kept on 3" x 5" cards. The report series and number and year are also coded. **DF9**
371. Dismuke, Nancy M., Irma S. Wachtel, and Katharine Way, "Proposal for Putting Nuclear Data on Punched Cards," Report No., ORNL-883 (Unclassified), Oak Ridge National Laboratory, Carbide and Carbon Chemicals Div., Oak Ridge, Tenn., 19 pp. (1951). A plan is proposed for coding certain nuclear properties and numerical values on McBee Keysort cards, so that all nuclei may be quickly searched for any of the coded data. Drawings of two proposed cards, one for a stable nucleus, and one for a radioactive nucleus, together with a detailed explanation of the suggested code are presented. **DF5**
372. Dix, W. S., Chairman, "Automation in the Library," ACRL Monograph, No. 17, 27-43, 50-51 (1956). Four papers presented at the 41st Conference of Eastern College Librarians two of which discuss the Univac and Remington Rand's 700 series of electronic data processing machines, the 3rd touches on machines and library design, and the 4th paper studies the possible applications of automation concepts to libraries. Bibliography. C.L.9:2. **JK11**
373. "Documentation of Spectra by Punched Cards," *Nature*, 176, 724, (15 October 1955). In order to make available infra-red and Raman data for use in physical or organic chemistry and for analysis, Butterworths Scientific Publications in London and Verlag Chemie have announced a joint Anglo-German scheme for providing the spectra of pure compounds and other technological products on punched cards together with other structural and spectral information. A.D. **J6 14**
374. Donnay, Gabrielle and J. D. H. Donnay, "A Punched Card Method for Computing Structure Factors," *Acta Cryst.*, 4, 74, 75 (1951). A method is described for calculating structure factors by punched cards, which is straight-forward and particularly rapid for crystals of low symmetry. C.A.45:4516. C.L.3:4. **J2 14**
375. Donnell, J. W., "If Computation Costs Too Much," *Chem. Eng.*, 58, No. 12, 138-41 (1951). Discusses punched card computing methods for engineering calculations. C.L.4:1. **J2 7**
376. Donnell, J. W. and Kenneth Turbin, "Multicomponent Vapor-Liquid Equilibrium Calculations Made by Punch-Card System", *Petroleum Refiner*, 29,

- No. 10, 109-11 (1950). Outlines a rapid and accurate calculation method which utilizes the standard IBM electric accounting machines. C.L.2:4, C.A.45:397. J2 7 14
377. Donohue, Jerry, "Systematic Calculation of Interplanar Spacings or Values of sine θ with Punched Cards (IBM)," *Acta Cryst.*, 3, 161-2 (1950). C.A.44:8192. J2 14
378. Dostert, L. E. (Georgetown University, Washington, D. C.), "Characteristics of Recent Mechanical Translation Experiments." Presented before the 126th national meeting of the American Chemical Society, New York, September, 1954. C.L.6:3. H17
379. Drillick, J. H., "New Electro-Mechanical System Provides Fast Access to Punched Card Data File," *Product Eng.*, 22, 176-78 (1951). Horizontal bars moving horizontally perpendicular to their length, offset all cards in file except one, by engaging notches in edge of card. A light beam passes through holes in cards and the data represented by the pattern of holes in the offset card may be interpreted by a system of shutters and a photocell. CK
380. Duer, M. D., and C. S. Lewis, "How we use IBM," *Library J.*, 78, 1288-1289 (August 1953). Description of the use of IBM equipment in the circulation and order departments of the University of Florida Libraries. Includes both routine operations and statistical studies. C.L.5:4. J11
381. Dunn, E. E., and G. E. Lynn, (Biochemical Research Department, The Dow Chemical Company, Midland, Michigan.) "Reporting and Indexing Biological Data by IBM Punched Card Methods." Presented at the 121st national meeting of the American Chemical Society, New York, March 1952. Indexing is obtained by numerical coding for individual chemical compound, chemical structure type, test organism, and test method. C.L.4:1. FJ6 12
382. Dyson, G. Malcolm, "Advances in Classification," *J. Document.*, 11, 12-18 (March 1955). Suggests a system for arriving at a multilateral classification of both topics and structures in the field of chemistry by devising machine language using self-demarcating code words suggested by Luhn. L.K. FOM6
383. Dyson, G. Malcolm, "Studies in Chemical Documentation III. Mechanized documentation," *Chemistry & Industry*, No. 16, 440-449 (1954). A brief description, with examples, of the I.U.P.A.C. provisional international system for the codification of organic compounds, the punching of the cards, and correlation and selection of information from the punched cards. C.L.6:4. L.K. FJM6
384. Dyson, G. Malcolm, "Advances in the Mechanical Documentation of Chemistry," *Chemistry & Industry*, 705 (1953). Discusses the available aids for using the chemical literature with maximum mobility and availability, with examples of problems which can be solved with modern mechanical aids. C.L.5:4. N1
385. Dyson, G. Malcolm, "The Preservation and Availability of Chemical Knowledge," *J. Chem. Educ.*, 29, 239-43 (1952). A system of coding chemical language is discussed as a means for recording and reproducing chemical information by machine. C.L.4:3. FJ6
386. "Eastern Joint Computer Conference, December 1954—Titles of Papers and Abstracts," *Computers and Automation*, 4, 15-17 (January 1955). L.K. K
387. Edge, Eleanor B., Norman G. Fisher, and Lucy A. Bannister (Chemical Department, Experimental Station, E. I. DuPont de Nemours & Co., Wilmington, Del.), "System for Indexing Research Reports Using a Punched-Card Machine." Presented at the 131st national meeting of the American Chemical

- Society, Miami, March 1957. A punched card system using the IBM 101 electronic statistical machine has been devised for indexing technical information. The index answers a myriad of detailed questions, ranging from the specific to generic, in the chemical field and in bordering sciences such as physics and biology. The system is designed to handle indexes, large numbers of chemical compounds and polymers. These are characterized by structural and compositional features rather than by name. Each compound and polymer component is indexed on a separate card along with types of reaction, properties, end uses, physical and biological studies, and techniques. C.L.9:1. JP9 15 16
388. "Electrofax Dry-photographic Enlarger," *J. Franklin Inst.*, **261**, 584-5 (May 1956) The RCA Electrofax machine is said to be the first enlarger printer designed for use with the Dexter developed Filmsort (punched-card) system for filing and selecting engineering drawings for reproduction. DL7 9
389. "Electronic Brain Translates Russian," *Chem. Eng. News*, **32**, 340-341 (1954). Describes how the electronic mathematical computer, IBM 701, has been converted into a language translator. K17
390. "Electronic Linquist: Cal Tech Scientist Turns Computer into Versatile Translator," *Wall Street J.*, **7** (June 28, 1957). Peter Toma of Cal Tech demonstrated his Datatron method for translating from four languages; French, Spanish, Russian and German. C.L.7:9. K17
391. Elersich, Valeria, (The Standard Oil Company (Ohio), Chemical and Physical Research Division, Cleveland 6, Ohio). "Abstracting of Petroleum Patents on (edge-) Punched Cards." Presented at the 122nd national meeting of the American Chemical Society, Atlantic City, September 1952. Coded on each card are product or process, inorganic and organic types of compounds involved in the invention, patentee, assignee, approximate date, also whether the novelty of the invention lies in the chemical reactants, the process operating variables, or the apparatus. C.L.4:3. D4 6 13
392. Fairbanks, E. E., *Econ. Geol.*, **41**, 761 (1946). Card system for the identification of ore minerals. 2 16
393. Fairthorne, R. A., "Notes on the NLL Card Catalogue of Aerodynamic Measurements," *J. Document.*, **10**, No. 1, 11-18 (1954). J14
394. Fairthorne, R. A., "Automata and Information," *J. Document.*, **8**, 164-172 (September 1952). Explains in non-technical language the operation of various types of automatic machines. C.L.5:1.L.K. JKN
395. Fairthorne, R. A., "Matching of Operational Languages in Documentary Systems," Library Memorandum No. 27. Royal Aircraft Establishment, Farnborough, England, 11 pp (1956). The principles underlying the development and use of scripts for information retrieval by machinery are considered. C.L.9:2. O9
396. "Faster than 300 Secretaries", *Business Week*, No. 1290, 108, 110, 112 (May 22, 1954). Shepard Laboratories, a New Jersey Company, has developed a typewriter that can cope with the output of an electronic computer. It is fed by punch cards or tape. C.L.6:3. K
397. "Filing Negatives and Transparencies," Eastman Kodak Co., Rochester, N. Y. (1953). Classification, filing and indexing systems are discussed. An edge punched card with attached film is illustrated. DLO
398. Fleisher, Harold, "An Introduction to the Theory of Information," *The Library Quarterly*, **25**, 326-32 (October 1955). An informative and easily followed discussion of the principles of elementary information theory. C.L.8:1. Q
399. Fleisher, Michael, (U. S. Geological Survey, Washington 25, D. C.). "Experi-

- ences with a Notched Card File of Geochemical Data." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. Literature reference on edge punched card (3500) subject coded. C.L.7:3. **D1 16**
400. Fletcher, J. H., and D. S. Dubbs, "Quick Access to Research Records," *Chem. Eng. News*, **34**, 5888-91 (1956). Cyanamid uses a molecular formula index in the form of a card file to record 35,000 organic compounds so that any one or related ones can be located in minutes. C.L.9:2. **A6**
401. Ford, Robert T., (Research Division, Sharp & Dohme Div., Merck & Co., Inc., West Point, Pa.) "Machine Correlation of Chemical and Biological Information." Presented at the 128th national meeting of the American Chemical Society, New York, September 1954. By assigning randomlike numbers to chemical components and superimposing those numbers in single fields, it is possible to express chemical structures on a single punched card. The National Research Council method for coding chemical structure has been used and sufficient room remains on the card so that basic pharmacological screening results also may be recorded. C.L.6:3. **FJ2 6 12**
402. "Fosdic II—Reads Microfilmed Punched Cards," *Nat. Bur. Standards (U. S.) Tech. News Bull.*, **41**, 72-74 (1957). The design and operation of Fosdic II (Film Optical Scanning Device for Input to Computers), a highspeed electronic device that can read microfilmed copies of punched cards and search for cards containing specific information, are outlined. C.L.9:3. **KL**
403. Foster, Laurence S. (Ordnance Materials Research Office, Watertown Arsenal, Watertown, Mass.). "Revision of the ASM-SLA Code for Classification of Metallurgical Literature." Presented at the 131st national meeting of the American Chemical Society, Miami, April 1957. Can be used in conventional or punched card filing systems. C.L.9:1. **FO1 7.**
404. Francisco, R. L., "Use of the Uniterm Coordinate Indexing System in a large Industrial Concern," *Spec. Libr.*, **47**, 117-23 (1956). A discussion of the Uniterm System for cataloging now in use at the General Electric Company Technical Data Center. C.L.8:3. **G**
405. Friedenstien, Hanna and Madeline M. Berry, "Chemical Documentation," *Reviews of Pure and Appl. Chem.*, **5**, No. 2, 109-12 (June 1955). Describes breadth and scope of chemical documentation. Discusses classification, cataloging, indexing, abstracting, reviews, searching, new tools—punched cards, etc.—"machine language" and need for terminology for various degrees of specificity and for generic relationships. **NO1 4 6**
406. Friedman, B. D., "Punched Card Primer," Public Administration Service, Chicago, 77 pp (1956). Functions, machines and techniques of IBM, Remington Rand, and Underwood punched card systems are described. C.L.9:3. **EJN**
407. Frome, Julius, and Jacob Leibowitz, "A Punched Card (IBM) System for Searching Steroid Compounds," U. S. Patent Office Research and Development Rept. No. 7 (July 1957). C.L.9:4. **JP6**
408. Frome, Julius, H. R. Koller, Jacob Leibowitz, and H. Pfeffer, "Recent Advances in Patent Office Searching—Steroid Compounds." Andrews, Dan. D., "—ILAS (The Integrated Logic Accumulating Scanner)," Patent Office Research and Development Reports, No. 8, Office of Research and Development, U. S. Patent Office, Washington, 15 pp. (1957). Describes an experiment in topological coding of chemical structure for storage and retrieval on SEAC at the

- N.B.S., and a proposed system for use with ILAS, using standard IBM cards.
FHK6 9
409. Fuchs, Otto, "Computers for the Solution of Chemical and Engineering Problems," *Chem. Ing. Tech.*, 25, 377-85 (1953). C.A.47:10285. K2 7
410. Gage, Robert P., "How Punch Cards Aid in Medical Research", *J. Am. Hospital Assoc.* (Sept. 1, 1956). Data about a patient is coded on an IBM card, and also written on the card between the rows of punches. "The Use of Punch Cards in Medical Research." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3. J2 12
411. Garfield, Eugene, "The Preparation of Printed Indexes by Automatic Punched-Card Equipment—A Manual of Procedures," Mimeographed report (24, March 1953), Medical Indexing Project, Johns Hopkins U., Baltimore 5, Md. A.D. JO11
412. Garfield, Eugene, "The Preparation of Subject-Heading Lists by Automatic Punched-Card Techniques," *J. Document.*, 10, 1-10 (March 1954). C.L.7:1. JO11
413. Garfield, Eugene, "Preliminary Report on the mechanical analysis of information by use of the 101 statistical punch card machines," *Am. Doc.*, 5, 7-12 (January 1954). Reviews the reasons why the use of machines is necessary for information analysis. Describes the major difficulties in using standard punched card equipment for information analysis. Describes the versatility of the 101 punched card machine. L.K. JO
414. Garrett, G. T., and O. Osmon, Jr., "New Punched Card System will Help you Organize Corrosion Data," *Chem. Eng.*, 64, No. 6, 342, 344, 346, 348 (1957). Edge-punched cards 5 x 7 in. used by American Potash and Chemical Corp., Trona, Calif. Describes simple and effective code under headings: A. Plant area; B. Materials; C. Corrosion Media; D. Conditions. DF7
415. Garrido, Jules, "Determination of Chemical Substances by Crystallographic Properties," *Bull. soc. franc. mineral*, 77, 989-95 (1954). X-Ray diffraction data and morphologic data on IBM cards. C.A.49:3605. J3 14
416. Garrott, P. B., "New Coding Systems Broaden Data Processing," *Automation*, 3, 70-76 (January 1956). C.L.8:3. F
417. Gey, Karl Friedrich, Hans Kalbe, Harold Schon, and Herman Stegemann, "Documentation of Physiological-Chemical Literature on Punched Cards," *Hoppe-Seyler's Z. physiol. Chem.*, 301, no. 1/2, 70-77 (1955). A direct coding system for use in the classification of physiological-chemical literature on hand sorted punched cards. A.D.7:4. DF1 6 12
418. Gilbert, Paul T., Jr., (Beckman Instruments, Inc., South Pasadena, Calif.) "An Optimal Punched Card Code for General Files." Presented before the 123rd national meeting of the American Chemical Society, Los Angeles, March 1953. An improved single-field superimposed coding system employs the spelling of words (descriptors) to be encoded, pairs of consecutive letters are selected from the descriptor to form the code. A modification of Zato-coding. C.L.5:1. DFP
419. Glazer, H. (Management Sciences Touche, Niven, Bailey and Smart, New York, N. Y.). "Linear Programming for Mechanized Data Handling." Presented before the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. FJ
420. "Glossary of Terms in the Field of Computers and Automation," *Computers and Automation*, 3, No. 10, 8-23 (December 1954). A.D.6:2. JKM

421. Gordon, A. H., "Adaptation of Mechanical Sorting and Tabulating Machines to Research in Marine Meteorology," *Meteorol. Mag.*, **80**, 269-70 (1951). J2 16
422. Gordon, A. H., "Development of Modern Techniques in Marine Meteorology," *Meteorol. Mag.*, **80**, 78-83 (1951). Meteorological observations punched into Hollerith cards. J2 16
423. Gould, Sydney W., "Permanent Numbers to Supplement the Binomial System of Nomenclature," *Am. Scientist*, **42**, 269-274 (April 1954). Advocates the use of numbers in plant classification as a permanent supplement to the binomial name for the species as now used. The use of machine-punched cards is basic to effective use of the proposed system. A.D.6:2. FJMO
424. Gradine, Joseph D., "Improved Structure Factor Computations on IBM Punched-Card Equipment," American Crystallographic Assn. Meeting, Chicago (1951). J2 14
425. Graham, M. H., B. A. Hildenbrand, and B. H. Weil, (Information Services Division, Ethyl Corp., Research Laboratories, Detroit, Mich.) "Indexing and Correlation of Fuel and Lubricant Additives by Machine-Sorted Punched Cards." Presented at the 129th national meeting of the American Chemical Society, Dallas, April 1956. C.L.8:1. JO2 6 7
426. Green, Robert S., "Welding Patent Classification in the A. F. Davis Welding Library," Ohio State University Studies, Engineering Series, Engineering Experiment Station Bulletin No. 140, Columbus, 74 pp (1950). Describes a simple yet comprehensive classification scheme for 12,000 patents, which are filed by patent number. An edge punched card is coded for each patent, giving classification, date of issue, and author or assignee. The illustration and abstract are pasted on the card. DO7 13
427. Green, John W., "The Use of Punched Cards in the Teaching of Qualitative Organic Analysis," *J. Chem. Educ.*, **28**, 638-40 (1951). Punched cards are used for recording and cataloguing properties of organic compounds dispensed as unknowns. C.L.4:1. D3 6
428. Greenhalgh, D. M. S., "Some new punched-card methods of Fourier synthesis," *Proc. Leeds Phil. Lit. Soc.*, Sect. 5, 301-7 (1950) cf C.A.45:949g. C.A.46:5959. J2
429. Greenhalgh, D. M. S., and G. A. Jeffrey, "New Punched-Card Method of Fourier Synthesis," *Acta Cryst.*, **3**, 311, 312 (1950). C.A.45:949. cf. C.A.46:5959. C.L.3:4. J2
430. Gronvik, Anna, "Modern Aids in Documentation," *Paperi ja Puu* (Paper and Timber), **38**, 475-77 (1956). (In Swedish). The Use of photocopying, micro-duplicating, and other reproduction methods and of punched cards and machine selectors, as aids to research and technical documentation, is surveyed. C.L.9:1. JLN1
431. Gruber, Wolfgang, "Der Stand der Verschlüsselung für die mechanische Selektion in der Organischen Chemie," *Dokumentation*, **I**, No. 9, 178-180 (November 1954). The author examines the possibilities of the Dyson and Wiswesser chemical notation systems as well as the possibilities offered by these methods with the use of punched cards. A.D.6:3. JM
432. Gruber, W., "The Use of Index Cards, Which are Punched in the Margin, for the Documentation of Organic Compounds," *Angew. Chem.*, **65**, 230 (1953). The author discusses the marginally punched card method and results obtained (in German). C.L.5:3. D6
433. Guibert, J., "Creation of an analytical punch card index of hydrocarbon de-

- posits," *Rev. inst. franc. petrole et Ann. combustibles liquides*, 9, 18-30 (1954) C.A.48:6104. J6
434. "Guide to NACE Corrosion Abstract Punch Card System," National Association of Corrosion Engineers, Houston. Edge punched cards. Gives classification scheme, items coded by random numbers; alphabetically arranged index of classified items; Code for journal references. DFO7
435. Gull, C. D., "Seven Years of Work on the Organization of Materials in the Special Library," *Am. Doc.*, 7, 320-29 (1956). The author discusses his experience with classification, subject headings, and coordinate indexing for controlling information. C.L.9:2. GNO
436. Gull, C. D., "Implications for the Storage and Retrieval of Knowledge," *The Library Quarterly*, 25, 333 (October 1955). A review of the various techniques used for storage and retrieval of knowledge from man's earliest days to the present date. N9
437. Gull, C. D., "Posting for the Uniterm System of Coordinate Indexing," *Am. Doc.*, 7, No. 1, 9-21 (1956). Ten manual and semiautomatic methods are considered in detail. The methods are given in a table and their relative efficiency and cost are described. C.L.8:2. GP8
438. Gull, C. D., "Instrumentation (in U. S. Government libraries)," *Library Trends*, 1, 103-126 (July 1953). Includes discussion of photocopying services, punched cards, the Rapid Selector, cataloguing codes and classifications. L.K. FJKO11
439. Guthrie, V. B., "Project 44—in the Cards," *Petroleum Processing*, 7, 1769-71 (1952). IBM cards in handling fundamental hydrocarbon data. C.L.5:1. J6
440. Gutenmakher, L. I., "Problem of Machine Technique in Scientific Information," *Vestnik Akad. Nauk S.S.S.R.*, No. 8, 46-52 (1952). This article is mainly of interest to show the similarity between American and Soviet approaches to the problem of scientific information. L.K. JK1 16
441. Hale, A. H., and J. W. Stillman, "Development of an Efficient Analytical-record System," *Anal. Chem.*, 24, 143-9 (1952). A system is described and illustrated which uses special forms, including edge punched cards. C.A.46:6877. DP3
442. Hangen, Welles, "Soviet Electronic Brain Equals Best in U. S., Americans Find," *New York Times*, p. 1 (December 11, 1955). A.D.7:2. K
443. Hardkopf, J. C., "Cybernetics and the Library," *Library J.*, 76, 999-1001 (June 15, 1951). Many tasks performed in the library are highly repetitive; therefore, many of them can be improved by applying a mechanical device. L.K. JK11
444. Harned, Jesse L., "The Practical Application of the Punched Card System in Assembling Statistical Data from the Medical Records," *Bull. Am. Assoc. Medical Record Librarians* (December 1940). J2 12
445. Hengstenberg, Otto, "Punched-Card Evaluation of Technical Data," *Stahl u. Eisen*, 71, 776-85 (1951). Examples from the steel industry are described. C.L.4:1. J2 8
446. Heumann, Karl F., "Data Processing for Scientists," *Science*, 124, 773-77 (1956). Survey of integrated and electronic data processing touching on the origin of the concepts, their use in business, machines that are available, indexing problems, and some scientific uses. C.L.9:2. JKO 8 16
447. Higginson, H. L., and A. Poplawska (Patent Office, Commonwealth of Australia, Canberra, A.C.T.), "Patent Specification Machine Searching the Field of Organic Chemistry." Presented at the 129th national meeting of the American Chemical Society, Dallas, April 1955. C.L.8:1. J6 9 13
448. Hill, Thomas T., "Finding Photographic Information," *J. Biol. Phot. Assoc.*,

- 17, No. 3, 103-14 (1949). Describes and illustrates coding, punching and use of edge punched cards for bibliography of photographic information. **DFL1**
449. "History of Tabulating Machines," *The Punched Card*, 1, 6-14 (1952-1953). Includes many illustrations and chronology 1780-1951. **JN**
450. Hoffman, Erwin F., "Multi-Dimensional Classification and the Transcription of Cancer Literature to Punched Cards", *Am. Doc.*, 3, 61-70 (1952). C.L.4:3. **DJ1 12**
451. Hodgson, M. L., C. J. B. Clews, and W. Cochran, "A Punch-card Modification of the Beevers-lipson Method of Fourier Synthesis," *Acta Cryst.*, 2, 113-16 (1949). C.A.43:6509. **J2**
452. Hollander, Gerhard L., "Bibliography on Data Storage and Recording," *Commun. and Electronics*, 49-58 (March 1954). Contains 330 titles and abstracts applicable to fields of data storage, recording, analogue-to-digital conversion, data presentation, and telemetering. C.L.7:1. **JK1 9**
453. Holmquest, Harold J. Jr., "Paleontological Identification and Analysis by the (edge-) Punched-Card Method," *Science*, 120, 897-898 (1954). A.D.6:1. **D2 16**
454. Holmstrom, J. E., "The Relation Between Reference Symbols and Language," *Rev. document.*, 17, 20-27 (1950). A.D.2:1. **FM17**
455. Hood, S. L., R. A. Monroe, and W. J. Visek, "Edge-Punched Cards for Scientific Literature References," ORO-102, U. S. Atomic Energy Commission, Oak Ridge, Tenn., 19 pp. ii (1953). Illustrates E-Z Sort card with 3 rows of holes along top and bottom and three fields having 23-word alphabet, two of which fields are for subject coding. **D1 9**
456. "How to Ferret Out Information Electronically," *Research Rev., Office Naval Research*, 9-16 (1956). The EDIAC, an electronic system on which Documentation, Inc., has carried out research for the Office of Naval Research, is described. C.L.9:2. **HK9**
457. Hudgens, C. R., and A. M. Ross, "Computing Fourier Syntheses in X-ray Crystal-structure Analysis—Improved Punched-card Method," *Anal. Chem.*, 25, 734-6 (1953). C.A.47:11963. **J2 14**
458. Hughes, E. W., "Punched-card Methods in Crystal-structure Calculations," Computing Methods and the Phase Problem in X-ray Cryst. Anal., X-ray Cryst. Anal. Lab., Dept. Phys., Penna. State Coll. 141-7, 1952. C.A.48:3748. **J2 14**
459. Hunt, Raymond, "Measuring the Utilization of Punched Card Machines," *Office Management and Equipment*, 13, No. 9, 20-23, 96-98 (1952). Discusses a method introduced by Prudential Insurance Company for determining the proportionate cost of each job accomplished with a punched card installation. C.L.4:4. **J8**
460. Hunter, T., and Graham M. Clark, "Electronic Data-Processing Machines," *Instruments and Automation*, 28, 782-93 (1955). From the simple punched card machine to the powerful electronic calculator. C.L.7:3. **JKN**
461. Hurd, C. C., "Electronic Data-Processing Machines," *Chem. Eng. Progr. Symposium Series No. 8*, Van Antwerpen, editor, 49, 49, American Institute of Chemical Engineers, New York (1953). Brief, general discussion and review of electronic machines for computation. **KN2**
462. Hyslop, Marjorie R., "Documentalists Consider Machine Techniques," *Spec. Libr.*, 44, 196-198, (May-June 1953). Report of the symposium on "Machine Techniques in Scientific Documentation" held at the Welch Medical Library of Johns Hopkins University on March 3, 1953. L.K. **JKN1**

463. Hyslop, Marjorie R., "International Classification for Metallurgical Literature," *Spec. Libr.*, **43**, 28-29 (January 1952). L.K. O7
464. "IBM Punched-Card Accounting is Adapted to Make Scholarly Indexes," *Publisher's Weekly*, 170, No. 19, 2150-2152 (1956). IBM is now producing the manuscript for a multiindex concordance to the works of St. Thomas Aquinas and has nearly completed a concordance of five of the Dead Sea Scrolls. A.D.8:2. J17
465. Imbrie, Margaret W., "Activities of the Literature Group in a Chemical Library," *J. Chem. Educ.*, **33**, 521-3 (1956). Includes obtaining and encoding data for an "end-use" file on 8" x 10" edge punched cards. DF4 6 9
466. Isbell, A. F., (Buckman Laboratories, Inc. Memphis, Tennessee), "An Improved Punch-Card System for Handling Scientific Information." Presented at the 120th national meeting of the American Chemical Society, New York, September 1951. A multiple direct code combination system, similar to the four-field, four-letter code system developed by Wise. C.L.3:3. DF9 16
467. Isbell, Horace S., "System for Classification of Structurally Related Carbohydrates," *J. Research Natl. Bur. Standards*, **57**, 171-78 (1956). A code number defines structure and configuration. Compounds can be selected by inspection of the code numbers or by punched cards. C.L.9:1. DFJ6
468. Jaffé, Hans H. (Department of Chemistry, University of Cincinnati, Cincinnati, Ohio), "Coding of Hammett Rho Values on (edge) Punched Cards," Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3. DF14
469. Jahoda, Gerald (Colgate-Palmolive Co., Jersey City, N. J.), "Uniterm Coordinate Indexing of Research Files." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3. G9 16
470. Jamieson, D. R., "Mechanized Bibliographical Aid," *Library Assoc. Record*, **53**, 216-321 (July 1951). Surveys mechanical apparatus for bibliographical searching, including the Rapid Selector, the electronic punched-card selector designed by Dr. Jacques Samain, various electronic computers, and the Ultrafax. L.K. JKN1 9
471. Jeffery, C. N., "Applications of Punched Cards to Patent Searching," *J. Inst. Engrs., (Australia)*, **26**, No. 6, 107-10 (1954). J9 13
472. Jellinek, E. N., Vera Efron, and Mark Keller, "Abstract Archive of the Alcohol Literature," *Quart. J. Studies Alc.*, **8**, 580-608 (1948); cf. *ibid.* **2**, 408 (1941). Punched card system. C.A.42:3226. D1 6 9
473. Jones, William S., and Peter H. Butterfield (Carbide and Carbon Chemicals Co., South Charleston, W. Va.), "A Technical Information Service Using (IBM) Punched Cards for Indexing and Retrieval." Presented before the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3. J9 16
474. Jones, Morton E., and Verner Schomaker, "Use of Punched Cards in Molecular-structure Determination. IV. Approximations to the Temperature Factor," *J. Chem. Phys.*, **19**, 511-12 (1951); cf. C.A.44:3349a. C.A.45:6441. J2 14
475. Kartha, Gopinath, "Double Fourier-synthesis. Punched card method," *J. Indian Inst. Sci.*, **35A**, 332-8 (1953). C.A.49:3652. J2
476. Keller, Mark, and Vera Efron, "The Classified Abstract Archive of the Alcohol Literature. I. Description of the Archive," *Quart. J. Studies Alc.*, **14**, 263-84 (1953). Hand sorted punched cards. C.A.47:10915. D1 6

477. Kemeny, John G., "Man Viewed as a Machine," *Sci. American*, **192**, 58-67 (April 1955). Discussion of electronic computers, a logical machine, the Turing machine, a suggested "universal" machine, and a reproducing machine. Only when we acquire a better understanding of the brain's amazing ability to call forth information will we be able to give a machine anything more than a limited memory. A.D.6:3. **KNQ**
478. Kendall, C. E., "Indexing of Data on Ultraviolet Absorption Spectroscopy," *Appl. Spectroscopy*, **9**, 158-160, 162-165 (December 1955). Lists six main requirements for an indexing system and offers the opinion that punched cards would appear to supply the only means currently available for the multi-parameter searches needed. A table indicating coding for edge punched cards is included. A.D.7:2. C.L.8:2. **DFO9 14**
479. Kendrew, J. C., "Use of a Computing Machine as a Mechanical Dictionary," *Nature*, **176**, 984 (19 November 1955). Discusses coding of edge punched cards. A.D.7:2. **DF**
480. King, Gilbert W., "The Asymmetric Rotor. IV. An analysis of the 8.5- μ band of D₂O by punched-card techniques," *J. Chem. Phys.*, **15**, 85-8 (1947); cf. C.A.40:30535. C.A.41:2333. **J2 14**
481. King, Gilbert W., George W. Brown, and Louis N. Ridenour, "Photographic Techniques for Information Storage," *Proc of the I.R.E.*, **41**, 1421-1428 (October 1953). Photographic media are examined as vehicles for the storage of digital information. A.D.6:2. **L9 16**
482. King, Gilbert W., and R. M. Hainer, "An Analysis of the 2.6- μ band of Hydrogen Sulfide by Punched-card Methods," *Phys. Rev.*, **74**, 1247 (1948). C.A.44:6269. **J2 14**
483. King, W. H., Jr. and William Priestley, Jr., "Spectrometric Analysis Employing Punch Card Calculators," *Anal. Chem.*, **23**, 1418-21 (1951). C.A.46:8565. **J2 14**
484. Kirschner, Stanley, "A Simple Rapid System of Coding and Abstracing Chemical Literature Using Machine-Sorted Punched Cards," *J. Chem. Educ.*, **34**, 403-5, (1957). **FJ4 6 9**
485. Kitz, N. and B. Marchington, "A method of Fourier synthesis with a standard Hollerith Senior Rolling Total Tabulator," *Acta Cryst.*, **6**, 325-6 (1953) (in English). C.A.47:8451. **J2**
486. Knaplund, Paul, Paul Fullerton, and Eileen Ford, "Here are Three Ways to Use Punched-Card Equipment," *Oil Gas J.*, **52**, No. 9, 70-71, 73, 75 (1953). Automatic punched-card equipment is applied to data-reduction calculations encountered in the exploration for oil. C.L.5:4. **J2 7**
487. Krieger, F. J., and W. B. White, "A Simplified Method for Computing the Equilibrium Composition of Gaseous Systems," *J. Chem. Phys.*, **16**, 358-60 (1948). C.A.42:4011. **J2 14**
488. Krull, A. R., "(Edge) Punch Card System for the Petroleum Industry," *Petroleum Engr.*, **28**, No. 3, E-27-E-29, E-32, E-34 (1956). Elementary description of coding and sorting, used for references in the Petroleum History Project. List of subject headings, direct coded, used singly or in combination. C.L.8:2. **DF1 6**
489. Kuentzel, L. E., "New Codes for Hollerith-Type Punched Cards," *Anal. Chem.*, **23**, 1413-18 (1951). For correlating chem. structure with absorption data. Coded on the card are wave length of absorption peaks, certain org. chem. structure features and elements, m.p. or b.p., and a serial no. or literature reference to the spectrogram. C.A.46:5436. **FJ2 6 14**

490. "Labor Saving Devices and Techniques of the Future", *Med. Library Assoc. Bull.*, 41, 60-68 (1953). Includes discussion of edge punched cards and book storage. C.L.5:2. **DN11**
491. Landee, Franc A. (Computation Laboratory, Dow Chemical Co., Midland, Mich.), "Numerical Data Handling Machines." Presented at the 124th national meeting of the American Chemical Society, Chicago, September, 1953. C.L.5:3. **KN2**
492. Lanham, B. E., and J. Leibowitz, (Research and Development, United States Patent Office, Washington, D. C.), "Establishment of a Flexible Vocabulary for Machine Handling of Information." Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. **JO9**
493. Lanham, B. E., J. Leibowitz, and H. R. Koller, "Advances in Mechanization of Patent Searching", *J. Patent Office Soc.*, 38, 820-38 (1956). Describes the relationships to be dealt with in coding patents, and the interfix device system of coding chemical structures. C.L.9:3. **FJ6 9 13**
494. Lanham, B. E., J. Leibowitz, H. R. Koller, and H. Pfeffer, "Organization of Chemical Disclosures for Mechanized Retrieval," U. S. Patent Office Research and Development Rept. No. 5 (June 1957). C.L.9:4. **JO4 6 9.**
495. Lenihan, J. M. A., "Isotope Catalogue on (edge) Punched Cards", *Brit. J. Appl. Phys.*, 3, 29 (1952). **D5**
496. Lester, A. M., "The Edge Marking of Statistical Cards", *J. Am. Statistical Assoc.*, 44, 293-4 (1949). **B2**
497. Levin, P., "Tools and Methods for Searching the Chemical Literature: a Selective Bibliography," MSLS thesis, Drexel Institute of Technology, 41 pp. (1955). C.L.8:1. **1 4 6**
498. Levine, Norman, "A Punched Card System for Filing Parasitological Bibliography Cards," *J. Parasitol.*, 41, 343-52 (1955). Edge notched cards 3" x 5". **D1 9 12**
499. Lingenberg, Walter, "The Use of Punched Cards in Libraries," *Arb. Bibl.-Lehrinstitut des Landes Nordrhein-Westfalen*, (9) 85 p. Diagr. Bibliog. (1955). A detailed description of the uses, peculiarities, application and prices of punched cards and Hollerith cards, with special reference to American practices, and German requirements. A.D.7:3. **DJ11**
500. Loftus, Helen E., and Allen Kent, "Automation in the Library—An Annotated Bibliography," *Am. Doc.*, 7, 110-26 (1956). The bibliography is intended to be selective and not exhaustive and represents material that came to the attention of the compilers prior to July, 1955, 93 references with abstracts. **N1**
501. Low, Ward C., "Technical Publication, Abstracts on IBM Punched Cards. I," Technical Note No. 15, Contract AF 19 (122)-36, Upper Atmosphere Research Laboratory, Boston University, 24 pp. (1952). Suggests a method of coding bibliographic information on IBM cards, and gives a code for subjects and equipment in the field of upper atmosphere physics. **FJ1 14**
502. Lowe, Ruth K., "Additional Bibliographic Uses for Keysort Punch Cards," *Library J.*, 76, 196-99 (1951). Details for use of Keysort punched cards in preparation and photostatic duplication of bibliographic lists. C.L.3:1. **DL1**
503. Luhn, H. P., "Self-Demarcating Code Words" (2nd ed.), I.B.M. Corp., New York, 84 pp., xiii (1956). Tables of specially composed three- and four-letter words which can be assigned to each of 20,810 information terms. L.K. **F**
504. Luhn, H. P. (Engineering Laboratory, International Business Machines Corp.,

- Poughkeepsie, N. Y.). "Statistical Approach to Mechanized Literature Searching". Presented before the 131st national meeting of the American Chemical Society, Miami, April 1957. C.L.9:1. **F2 9**
505. Luhn, H. P., "A New Method of Recording and Searching Information," *Am. Docu.*, 4, 14-16 (January 1953). Method uses the principle of characterizing a topic by a set of identifying elements or criteria. C.L.5:2. **F9**
506. Lukens, H. R., Jr., E. E. Anderson, and L. J. Beaufait, Jr., "Punched Card System for Radioisotopes," *Anal. Chem.*, 26, No. 4, 651 (1954). Edge punched cards used to identify an isotope on the basis of its radiation characteristics and half life. **D5**
507. LuValle, James E., "Bibliography on Photographic Theory," Investigators Restricted Seminar No. 1 on the Chemistry of Photographic Processes. Sponsored by the Chemistry Division of the Headquarters Air Research and Development Command (RDRRC), Chicago, Sept. 4, 1953. **L1**
508. Luzzati, Victorio, "Mechanical Calculation of Structure Factors," *Bull. soc. franc. mineral. et. crist.*, 73, 601-3 (1950). C.A.45:3678. **J2 14**
509. MacCasland, G. E., "A Concise Form for Scientific Literature Citations," *Science*, 120, 150-152 (1954). Proposes a form of concise literature citation utilizing numbers for year and page and capital letters as code for the journal name which could be adapted to automatic sorting devices for punched cards and microcards. L.K. **F1**
510. McCrone, W. C., "Punched-Card System for Tabulation of Crystallographic Data," *Anal. Chem.*, 28, 972-75 (1956). A punched card system is proposed for collecting and using crystallographic data. Enough data are punched on the cards to permit all quantitative crystallographic properties to be read directly or calculated from punched data. C.L.9:2. **JP2 14**
511. McCrone, Walter C., Jr., "Classification of Analytical Methods," *The Frontier*, 9, No. 4, 9-11 (1946). **O3**
512. MacDonald, K. C., "Information Theory and Its Application to Taxonomy," *J. Appl. Phy.*, 23, 529-31 (1952). Information theory is applied to the problem of classification of data, and several models are discussed which represent various possible methods of filing data with the purpose of determining the optimum size of filing-unit in relation to the given data. C.L.5:1. C.A.46:8429. **OQ9**
513. McGaw, Howard F., "Marginal Punched Cards in College and Research Libraries," The Scarecrow Press, Washington, D. C., 218 pp., (1952). Listed in *Chem. Eng. News*, 30, 2537 (1952). C.L.4:3. **D1 11**
514. "Machine Age in the Library", *Chem. Week*, 74, No. 11, 74, 76 (1954). Describes development being carried on at Battelle Memorial Institute for Technical information encoding and retrieval using IBM punched cards and machines. C.L.6:3. **J9**
515. Magalhaes, Hulda, "The Golden Hamster as a Laboratory Animal", *J. Animal Technicians Assoc.*, 5, No. 2. Edge punched cards with single row of holes used for bibliographic references. Seventeen "Primary Subjects" direct coded. Twenty "Secondary Subjects" under each "Primary Subject" in a twenty-position direct coded field. **DF1 12**
516. Maginnity, Paul M., and Nancy J. Allison (Callery Chemical Co., Callery, Pa.), "Use of Machine-Sorted Punched Cards in Literature of Boron Chemistry," Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. **J1 4**

517. Maierson, Alvin T., and W. W. Howell, "Application of Standard Business Machine Punched-Card Equipment to Metallurgical Literature References," *Am. Doc.*, 4, 3-13 (1953). C.L.5:2. J1 7
518. Marsh, J. L. (Ciba Pharmaceutical Products, Summit, N. J.), "An Index of Chemical Structure" (on machine punched cards). Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3. J4 6
519. Masuyama, M., "An Elementary Method of Construction of Punched Cards for p^a-and Other Designs," *Repts. Statist. Appli. Research, Union Japan Scientists and Engrs.*, 4, 78-84 (1956). (In English). A deck of either hand- or machine-sorted punched cards can be used to construct factorial, incomplete block, lattice, and other experimental designs. DJ2
520. Mathay, W. L. and R. B. Hoxeng, "A Classification and Filing System for Corrosion Literature," *Corrosion*, 12, 588-92 (1956). The system involves an adaptation of the NACE subject filing index for use with NACE corrosion-abstract punched cards. C.L.9:1 DFO7
521. "Mechanization in Libraries," *Library Trends*, 5, No. 2 (1956). Entire issue devoted to ten articles on this topic. C.L.9:1. DGJN11
522. "Mechanized System Launches New Era for Literature Searching," *Chem. Eng. News*, 30, 2806-8, 10 (1952). A new system of handling information is based on the standard IBM card but depends upon photoelectric scanning. C.L.4:4. EK9
523. "Mechanical Aid for Sales Brain," *Chem. Week*, 74, No. 22, 52-54 (1954). Chemical sales analyzed and tabulated by machine punched cards. C.L.6:3. J8
524. "Mechanized Copying, Filing Cuts Processing, Handling Time," *Ind. Labs.*, 7, No. 1, 59 (1956). Microfilm inserts in punched cards (Film-sort aperture card devised by Dexter Folder Co., Pearl River, N. Y.) are used by the government for finding, viewing, reproducing, handling and refiling of engineering drawings and related documents. C.L.8:1. JL7 9
525. "For the Memories", *Ind. Bull. of Arthur D. Little, Inc.* No. 269, 2 (1950). Discusses the principles and uses of memory machines. C.L.2:4. KN
526. "A Method of Coding Chemicals for Correlation and Classification," Chemical-Biological Coordination Center, National Research Council, Washington 25, D. C., 98 pp., (1950). Rules and directions for use of a "code devised primarily to permit the use of punched cards in the correlation of chemical structure with biological action". C.L.3:1. FJ2 4 6 12
527. "Military Specifications for Microfilming Engineering Drawings and Related Data," Bureau of Aeronautics, Dept. of the Navy, MIL-M-18872 (Aer) 13 June, 1955. L7
528. Mitchell, H. F., "The Use of the UNIVAC Fac-Tronic System in the Library Reference Field," *Am. Doc.*, 4, 16-17, (January 1953). Results of the study indicate that this equipment could be satisfactorily adapted to the library reference situation, however, as the UNIVAC is now constructed it is too large a capital investment for any but the very large libraries. L.K. HK1 11
529. Mohrhardt, F. E. (Library, United States Department of Agriculture, Washington, D. C.), "Critique on the Development of Mechanized Data Handling", Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. N
530. Mooers, Calvin N., "Making Information Retrieval Pay," *Zator Technical Bulletin* No. 553, Zator Company, Boston, 13 pp. (1950). (Presented at the

- 118th national meeting of the American Chemical Society, Chicago, September 1950.) C.L.2:3. **HKO9**
531. Mooers, Calvin N., "Coding, information retrieval and the Rapid Selector," *Am. Doc.*, 1, 225-229 (October 1950). A mathematical criticism of the code suggested for the Rapid Selector by Wise and Perry. (no. 275, first ed. biblio.) L.K. See No. 668 **FKN2**
532. Mooers, C. N., "Zatocoding Applied to Mechanical Organization of Knowledge," *Am. Doc.*, 2, 20-32 (January 1951). Examines concepts inherent in information retrieval by machine and compares these concepts with the conventional methods of classification-indexing and card cataloging. Discusses the principles and practices of Zatocoding. L.K. **CFNO9**
533. Mooers, Calvin N., "Scientific Information Retrieval Systems for Machine Operation—Case Studies in Design", (Zator Technical Bulletin, No. 66, mimeographed) Zator Company, Boston, 1951. Presented at the 11th International Congress of Pure and Applied Chemistry, New York, September, 1951. The case studies presented make use of the Zatocoding System of information retrieval which employs an edge-notched card. A.D.4:3. **CHP9**
534. Mooers, Calvin N., "Choice and Coding in Information Retrieval Systems," *Transactions of the I.R.E.* 112-118 (1954). Information retrieval systems are susceptible to treatment by communication theory at the coding and machine level, and there are a number of analogies between retrieval systems and multiplex signalling systems. A.D.7:2. **FQ9**
535. Mooers, Calvin N., "Tabulation of Characteristics of Retrieval Systems," *Zator Technical Bulletin* No. 84a, Zator Company, Boston (1953). (presented at the annual meeting of the American Documentation Institute, November 6, 1953, Washington, D. C.). **N9**
536. Mooers, Calvin N., "Information Retrieval Viewed as Temporal Signalling" (presented at the International Congress of Mathematicians, Harvard University, Aug. 30 to Sept. 6, 1950). *Proceedings*, 1, 572. Points out that the process of signalling across a spacial span in communication has its analogue in signalling across a temporal span in retrieval. **Q9**
537. Mooers, Calvin N., "Zatocoding and Developments in Information Retrieval," *Aslib Proc.*, 8, no. 1, 3-22 (Feb. 1956). Describes the edge-notched cards, the selector device, the superimposed random coding techniques and establishment of descriptors, all of which comprise the Zatocoding system for information retrieval. A.D.8:1. **CFKOP9**
538. Mooers, Calvin N., "Information Retrieval on Structured Content." Presented at the Third London Symposium on Information Theory, Sponsored by the Department of Electrical Engineering of the Imperial College of Science and Technology and held at the Royal Institution, September 12-16, 1955. The cost of processing information in digital computers can be substantially reduced by more extensive intellectual preparation. This involves the creation of interlocking sets of "descriptors", their symbolic representation as "n-tuples," and subsequent conversion into structures susceptible of diagramming and manipulation. A.D.7:2. **KOQ**
539. Mooers, Calvin N., "The Zator—A Proposal, A Machine for Complete Documentation." *Zator Technical Bulletin*, No. 65, Zator Co., Boston, 20 pp. (1951). **CKP9**
540. Mooers, Calvin N., "Ciphering Chemical Formulas—The Zatopleg System," *Zator Technical Bulletin* No. 59, Zator Company, Boston, 8 pp. (1951). **F4 6**
541. Mooers, Calvin N., "Application of Random Codes to the Gathering of Statis-

- tical Information" (Master's thesis, Massachusetts Institute of Technology, Dept. of Mathematics, Feb. 1948) Published as Zator Technical Bulletin No. 31, 28 pp. (1949). **F2**
542. Mooers, Calvin N., "The Exact Distribution of the Number of Positions Marked in a Zatorcoding Field," Zator Technical Bulletin, No. 73, Zator Company, Boston, 9 pp. (1952). **FQ**
543. Morino, Yonezo, and Kozo Kuchitsu, "Computation in Electron-diffraction Investigation of Gaseous Molecules by means of Punched Card Machine," *X-Sen (X-Rays)*, 8, 37-43 (1954). C.A.49:7965. **J2 14**
544. Moyer, S. R., "Automatic Search of Library Documents," *Computers and Automation*, 6, 24-29 (May 1957). Describes the design of a system for storage and retrieval of documents using the IBM 705 electronic data processing machine. Modified coordinate indexing is used. C.L.9:3. **GKP9 11**
545. "Multidimensional Indexing Speeds Instrument-Document Search at NBS," *Chem. Processing*, 19, No. 4, 96-97 (1956). Each 5" x 8" card represents an index term, and the identity of a document to which that term applies is noted by punching a hole in the card at the appropriate location. C.L.8:3, **EO9**
546. "NACE Provides Punch Card Service", *Library Bull. Abstracts*, 26, No. 31, 124 (1951). The National Association of Corrosion Engineers has provided an abstract punched card system to aid in solving corrosion problems. C.L.3:4. **DP7 9**
547. Naimark, George M., "Industrial Analytical Record Keeping", *Drug & Cosmetic Ind.* (Sept., 1955). Describes and illustrates forms, including edge punched cards, and procedures developed in pharmaceutical company. Gives criteria for record systems. Compares methods before and after adopting this system. **DP3 12**
548. Naimark, George M. and Robert F. Prindle, "Pharmaceutical Control Laboratory Record System," *Anal. Chem.*, 26, 645-7 (1954). Two forms, a marginally punched permanent record card and an analytical work sheet, are the basis for a rapid and efficient pharmaceutical control lab. record system. C.A.48:10993. **DP3 12**
549. Neil, A. V., "Machines or Books—a Case for Both", *Stechert-Hafner Book News*, 9, 1-2 (January 1955). Machines and information coding have advanced beyond the stage of inventors' dreams. They are an actuality and librarians must prepare themselves for this new era in librarianship by learning new nomenclature and new techniques. L.K. **N11**
550. Nelson, W. L., "Technical Filing System," *Oil Gas J.*, 54, No. 25, 111 (October 24, 1955). The author argues that there are practical limits to an alphabetical file and therefore a more elaborate numbering system is justified. A numbering schedule for materials on petroleum is presented. C.L.8:1. **FQ6**
551. Newman, Simon M., "Problems in Mechanizing the Search in Examining Patent Applications," Office of Research and Development, U. S. Patent Office, U. S. Dept. of Commerce, Washington, 29 pp. (1956). Discusses problems in patent searching and examination with lengthy examples. No discussion of mechanization, discusses what must be encoded. **FN13**
552. Newman, Simon M., "Storage and Retrieval of Contents of Technical Literature, Non-chemical Information," Patent Office Research and Development Reports, No. 4, Office of Research and Development, U. S. Patent Office, Washington, 16 pp. (1957). Discusses research on linguistic method of encoding non-chemical scientific information for retrieval. **FH13 16 17**
553. Newman, Simon N., "Linguistic Problems in Mechanization of Patent Search-

- ing," Patent Office Research and Development Reports, No. 9, Office of Research and Development, U. S. Patent Office, Washington, 10 pp. (1957). Discussion of terminology and classes and their organization for coding. FMO13 17
554. Nowacki, W., "The Application of Punched-card Machines for Fourier and Patterson Syntheses of Crystals," *Chimia* (Switz.), 2, 274-5 (1948). C.A.43: 6509. J2 14
555. Ohrmund, Margaret (Eli Lilly and Co., Indianapolis, Indiana), "An Antibiotic Literature File for Chemists," Presented at the 126th national meeting of the American Chemical Society, New York, September 1954. Physical and chemical data coded on IBM cards. C.L.6:3. J1 6 12
556. Opler, Ascher (Research Dept., Western Division, The Dow Chemical Company, Pittsburg, California), "Contributions to the Theory of Automatic Information Retrieval." Presented at the 129th national meeting of the American Chemical Society, Dallas, April 1956. For automatic coding and subsequent machine retrieval of information, the language of an indexing reference is converted to its syntactic-topological equivalent and thence to a digital sequence suitable for storage and recovery by a high-speed, large-capacity computing machine. C.L.8:1. FHK09 17
557. Opler, Ascher, "A Topological Application of Computing Machines," AIEE Special Publication T-85, "Proceedings of the Western Joint Computer Conference" (Meeting 1956), p. 86 (1956). Describes a method of coding the topology of networks that contain elements used repeatedly in various arrangements, such as chemical structural formulas, electric circuits, engineering diagrams, traffic flow patterns. The code is reduced to a digital sequence amenable to storage in digital computers. FK7 16
558. Orosz, Gabor, "New Method for Document Retrieval by Means of Punched Cards," *Dokumentation (Leipzig)*, 2, 21-25 (March 1955) 2, 41-49 (May 1955). Detailed examination of the adaptability of the Hollerith system to document retrieval, including practical as well as theoretical considerations and concentration on the problem of coding with numerous examples for the variant forms. A.D.7:1. FJQ9
559. Orosz, G., and L. Takacs, "Some Probability Problems Concerning the Marking of Codes into the Superimposition Field," *J. Document.*, 12, 231-34 (1956). Exact formulas (based upon K. Jordans' probability theorem) are given for the distribution of the marked sites in a code field and for the distribution of multiple marking of a site. C.L.9:2. FQ2
560. Orr, Charles H., "A Punched-Card (edge punched) System for Graduate Research," *J. Chem. Educ.*, 30, 140-42 (1953). A system is described which is suitable for preparation and use by an individual student. Includes an interesting, yet simple numerical code. C.L.5:2. DFP1 4
561. Parker, I. M., "The Literature on Corrosion," *Corrosion*, 7, 450-54 (1951). The National Association of Corrosion Engineers adopted a program whereby abstracts from various sources are collected, classified, and published. C.L.4:2. I 7
562. Passer, Moses, "A New Photocopying Process for (edge) Punched Cards," *J. Chem. Educ.*, 33, 581-3 (1956). Eastman Verifax Printer, and Photostat Instant copier can copy from books (e.g. Chem. Abs.) and makes a positive print on nonsentive cards made from 110 lb. Hammermill Index. DL
563. Patterson, Louise, "A Punch-card Abstract File on Solid State and Transistor Literature," Mass. Inst. Tech. Lincoln Lab. Tech. Rept. no. 74, 48 pp. Lexington, Mass. (12 Jan 1955). A.D.6:4. F1 9 14

564. Patterson, Gordon D., Jr., and M. G. Mellon, "Classification of Methods in Quantitative Analysis," *J. Chem. Educ.*, **26**, 468 (1949). O3
565. Patton, A. R., "Punch Card Filing System for your Slides," *The Camera*, **73**, No. 1, 63, 130 (1950). Regularly spaced holes are punched along top and bottom of Kodachrome slides, which are then coded for the subject matter in the picture and sorted like edge punched cards. DL9
566. Pentz, E. Irene, and Ralph Bushee, "A Classification and Filing System for Scientific Reprints and References," *J. Chem. Educ.*, **32**, 622-3 (1955). A.D.7:3. O1 16
567. Perry, J. W., Allen Kent, and M. M. Berry, "Mechanized Literature Searching—A Progress Report," *College and Research Libraries*, **16**, 157-163 (April 1955). Discusses analysis and organization of subject matter for mechanical retrieval of technical information. O9
568. Perry, J. W., Allen Kent, and M. M. Berry, "Mechanical Literature Searching—A Progress Report," (Text of paper presented before the Technical Information Panel, Atomic Energy Commission, April 30, 1954). Battelle Memorial Institute, Columbus, Ohio, 10 pp. (1954). O9
569. Perry, J. W., and Allen Kent, "The Russians Have a Machine for It. A Review," *Am. Doc.*, **7**, 233-34 (1956). A description of the use of a Russian experimental searching machine which resembles a modified IBM sorter. C.L. KN
570. Perry, James, "Mechanical Documentation: Recent Advances and Applications," A lecture at Columbia University August 21, 1953. Mimeographed Circular, 21 pp. Experimental and tentative methods are described, use of semantic factoring and formation of a code dictionary for use with new IBM punched card machine designed for literature searching. FHKO9
571. Perry, J. W., M. M. Berry, F. U. Luehrs, and Allen Kent, "Automation of Information Retrieval," in Proceedings of the Eastern Joint Computer Conference, Philadelphia, December 8-10, 1954, pages 68-72, New York: The American Institute of Electrical Engineers, 1955. Advantageous application of computer-type equipment to automation of information retrieval is determined by ability to analyze the subject content of documents and records in terms of those aspects of meaning that are important in identifying information of pertinent interest. A.D.7:2. KO9
572. Perry, J. W., Allen Kent, and Madeline M. Berry, "Machine Literature Searching," Interscience Publishers, Inc., New York, 162 pp., 1956. Various methods and techniques for automatic searching of large information files. C.L.8:3. N9
573. Perry, James W., and Robert S. Casey, "Mechanized Searching," in "Encyclopedia of Chemical Technology," Kirk and Othmer, eds. 449-467, The Interscience Encyclopedia, Inc., New York: 1952. L.K. N9
574. Perry, J. W., "Specialized Machines of the Future for Handling Broad Ranges of Subject Matter," *Am. Doc.*, **4**, 22-26 (January 1953). L.K. K9
575. Pietsch, Erich, "On the Question of Coding in Mechanical Documentation," *Nachr. Dokument.* (Frankfurt am Main), **7**, no. 3, 117-119 (Sept. 1956). A.D.8:1. F1
576. Pietsch, Erich, "Mechanized Documentation in Industry," *Nachr. Dokument.*, **4**, 80-86 (June 1953), Methods in use in Germany are described, including the IBM card system of the Gmelin-Institut fur anorganische Chemie, and edge punched card systems. L.K. DJ4 8
577. Pietsch, Erich, "Mechanized Documentation: Significance in the Economics of Scholarly Activities," *Nachr. Dokumentation*, **3**, 3-5 (March 1952). C.L.4:3. N1 8 16
578. Pietsch, Erich, "New Control Methods for Exact Knowledge in Science and

- Technology," *Nachr. Dokument.*, 2, No. 2, 38-44 (1951). The Gmelin Institute of the Max Planck Gesellschaft is trying out some punched card techniques in handling its huge volume of literature and data in inorganic chemistry. C.L.4:1. **HJ 1 4**
579. Porter, Betty Brown, and George S. Crandall, "The Use of Machine-Sorted Punched Card in Documentation," Third World Petroleum Congress, Vol. X. Preprint 7 (Leyden, Holland) 1951. A method for maintaining a centralized library bibliography covering the petroleum field wherein searching is rapidly carried out by punched card machines of the IBM or Hollerith type is described. C.L.3:3. **J1 6**
580. Poulin, Roger, J., "A Punched Card Shelf List for Discarding Books," *Am. Doc.*, 7, 229-30 (1956). C.L.8:4. **D11**
581. "Preliminary Report on Research in Progress in Scientific Documentation", National Science Foundation, Office of Scientific Information, Washington, D. C. Mimeographed, 25 pp (Aug. 1956). Brief descriptions of research being done by each of 22 various groups and individuals in the areas: Organization of Information; Equipment for Storage and Retrieval; Special Studies; Mechanical Translation. **HKNO17**
582. "Punched Cards: Standards Committee Meets", *World's Paper Trade Rev.*, 147, 1433 (1957). A British committee and three subcommittees have been formed to consider standards for the shape, size, and content of punched cards. One of the subcommittees is concerned with punched tape. C.L.9:3. **H**
583. "The Punched Card" Vol. 1, The Punched Card Publishing Co., Detroit (1952-53). A business annual publication dedicated to the advancement and enlightenment of the science of machine punched card accounting. **N8**
584. Rabinow, Jacob, "The Notched Disk Memory", *Elec. Eng.*, 71, No. 8, 745-9 (1952). Information may be stored and rapidly retrieved through magnetic pulses on both sides of thin metal disks. Whole disk is scanned in one revolution and segment-like notch in disks permits magnetic scanning head to move rapidly from one disk to another. **K9**
585. Rabinow, J., "DOFL First Reader," Technical Report No. TR-128. Diamond Ordnance Fuze Laboratories, Ordnance Corps., Dept. of the Army, Washington, 39 pp. (Nov. 26, 1954). Character recognition device, light from image through perforated rotating scanning disc falls on photoelectric cell, discusses different methods of scanning. **K**
586. Randall, G. E., "Practicality of Coordinate Indexing," *College and Research Libraries*, 15, 417-419 (October 1954). Comments derived from an inspection of 45 ASTIA Uniterm Cards. A.D.6:1. (See no. 631.) **GN**
587. "Rapid Nuclide Identification and Selection by Nine Major Classifications," Circular, Radioactive Products, Inc., Detroit (1953). Describes commercially available file of edge-punched cards, radioisotopes classified and coded by, half-life, alpha, beta, and gamma energies, availability, mass and atomic numbers, and symbol. **D5**
588. Reed, Roger W., and Kenneth F. Gregory, "A Punch Card (edge-punched) for Abstracts of Bacteriological Papers," *Science*, 118, 360-1 (Sept. 25, 1953). **D1 12**
589. Rees, Mina, "Computers: 1954," *Sci. Monthly*, 79, 118-24 (August 1954). Included in this discussion of the scientific and industrial uses of computers is a section on mechanical translation. A.D.5:4. **K8 16 17**
590. Rees, Thomas H., Jr, and Allen Kent, "Jargon of Machine Literature Searching," (Center for Documentation and Communication Research, School of

- Library Science, Western Reserve University, Cleveland 6, Ohio). Presented at the 131st national meeting of the American Chemical Society, Miami, April 1957, C.L.9:1. **M9**
591. "Report by the Advisory Committee on Application of Machines to Patent Office Operations," U. S. Department of Commerce (Sales and Distribution Division, Room 6227, Commerce Building, Washington) 75 pp. (December 22, 1954). For improving information organization and retrieval, to facilitate patent searching. C.L.5:7. **NO9 13**
592. Reumuth, H., "The Indexing of Chemical Compounds. A Contribution to the Problem of Organization of the Literature," *Z. Angew. Chem.*, **41**, 1264-7 (1928). A special form of card index is described which, by means of a series of tabs, permits the subject matter to be classified in several ways at once. C.A.23:743. **BO4 6**
593. Rhynders, Robert W., "Product History Clears Haze from Technical Records," *Ind. Labs.*, **7**, No. 2, 30-1 (1956). Gives a suggested outline for preparation of the history of a product from its conception through successful production. Information and references are kept on edge punched cards. C.L.8:2. **D7 8**
594. Ridenour, Louis N., Editor, "Modern Physics for the Engineer," McGraw-Hill Book Company, New York, 1954. In Chapter 17, "Communication Theory and the Transmission of Information," J. B. Wiesner traces communication engineering in the course of its development. In Chapter 18, "Computing Machines and the Processing of Information", Dr. Ridenour gives a historical introduction to computers. A.D.6:2. **KNQ7 14**
595. Ridenour, L. N., "Bibliography in an Age of Science," University of Illinois Press, Urbana, pp. 5-35 (1951). After analyzing the function of a library, the author suggests that books should be discarded in favor of electronic means of storage of information for many purposes, and that an automatic cataloguing machine is possible. L.K. **HN1 11**
596. Rose, Arthur M., "The Application and Use of Punched Cards and International Business Machines in X-ray Crystal Structure Analysis," Univ. Microfilms (Ann Arbor, Mich.) Publ. No. 6006, 159 pp. microfilm \$1.99, paper enlargement \$15.90; *Dissertation Abstr.*, **13**, 1006-7 (1953). C.A.48:4308. **J2 14**
597. Rose, Arthur and Theodore J. Williams, "Punched Card Devices for Distillation Calculations," *Ind. Eng. Chem.*, **42**, 2494-7 (1950). C.A.45:1389. **J2 14**
598. Rose, Arthur and Theodore J. Williams (The Pennsylvania State College, State College, Pa.), "Micropunch Cards." Presented at the 121st national meeting of the American Chemical Society, New York, March 1952. Explanation and descriptions for preparation of IBM microcards, (1) from ordinary microfilm using conventional photographic techniques; (2) using ordinary microcard negatives and conventional photographic printing on available flat film. C.L.4:1. **JL**
599. Ruston, W. R., "Edge-notched Cards as Aids in Science Documentation," *Nachr. Dokument.*, **3**, 5-12 (March 1952). Applications of edge punched cards, with Universal Decimal Classification coding, are described and illustrated for metals and their compounds. C.L.4:3. **DFO4**
600. Sanford, J. A. and F. R. Theriault, "Problems in the Application of Uniterm Coordinate Indexing," *College and Research Libraries*, **17**, No. 1, 19-23 (1956). A report of a large-scale uniterm coordinate index covering 70,000 documents created for the library of the National Security Agency. The percentage of retrieval is stated as "very high". C.L.8:2. A.D.7:3. **G9 1 1**
601. Savedoff, Lydia, Jack Belzer and Herrick L. Johnston, "Computation of Ther-

- modynamic Properties from Spectra with the Aid of Punched Card Equipment," Third symposium on combustion and flame and explosion phenomena, Univ. of Wis., Madison, Sept. 7-11, (1948), 595-9 (1949). C.A.46:11626. J2 14
602. Sebring, M. W., "A Marginal Punched Card System for a Specialized Information Collection," *Am. Doc.*, 4, 18-22 (1953). Application of system to the field of abrasive products. C.L.5:2. DP7
603. Scheele, Martin, "Certain Fundamental Problems in Documentation by Means of Punched Cards," *Nachr. Dokument.*, 6, No. 3, 111-16 (Sept. 1955). Describes meaning and limitation of documentation, punched card system fundamentals, and its distinction from the classical methods. C.L.8:1. A.D.7:3. N1
604. Scheele, Martin, "Punched Cards in the Biological Sciences," Schweizerbart'sche Verlagsbuchhandlung Stuttgart: 114 pp., 1954. This introductory book on punched cards for managing biological research records covers the fundamentals of the Cordonnier system, edge punched cards, machine sorted cards, coding, and their application to research, and documentation. C.L.9:2. A.D.7:2. DEJ12
605. Schell, Emil D., (Chief, Mathematical Computation Branch, Department of the Air Force) "Project Scoop," *Systems for Modern Management*, 17, 7-8, 35 (May 1953). A description of the use of UNIVAC to compute military requirements. A.D.5:1. K2
606. Schlink, F. J., "Getting the Most Out of Index Cards," *Industrial Management*, 55, 135-8 (Feb. 1918). Ordinary file cards are notched in given positions along the edges. One-quarter inch spacing is suggested for qualitative data, and scaled edge data recorded to nearest 0.03 in. The file may be inspected and sorted manually or it may be sorted by appropriately spacing one or more rods in the bottom of the file drawer to separate the desired cards. C9
607. Schurmeyer, Walther, "Selecto: A New Selecting Technique for Documentation," *Nachr. Dokument.*, 3, 33-34 (March 1952). A modified "peek-a-boo" type of sorting is employed in Cordonnier's Selecto technique. C.L.4:3 E9
608. Sharpe, John A. and Paul W. Fullerton, "An Application of Punched Card Method in Geophysical Interpretations," *Geophysics*, 17, 707-20 (1952). Use of a trigonal coordinate system and substitution of finite sums for the infinite integrals leads to a particularly useful form of adequate practical accuracy for calculation by punch card methods. C.L.5:1. J2 14
609. Shaw, Ralph R., "Mechanical and Electronic Aids for Bibliography," *Library Trends*, 2, No. 4, 522-31 (1954). The basic processes which form the parts of any bibliographic system and the operations involved in each are outlined; their possibilities for mechanization are considered. C.L.6:4. N1
610. Shaw, Ralph R., "Will the Machines Take Over?" *Library J.*, 76, 1085-1087 (July 1951). Library work can be divided into two parts, one intellectual and the other mechanical. Discusses the importance of standardization to mechanical process. L.K. N11
611. Shaw, Ralph R., "The Rapid Selector," *AIBS Bulletin*, 4, 20-21 (July 1954). A.D.5:4. KL
612. Shaw, Ralph R., "High Speed Intermittent Camera," *Am. Doc.*, 1, 194-196 (October 1954). Describes the operation and disadvantages of the original Rapid Selector. Outlines improvements which have been made in the equipment. L.K. KL
613. Shaw, Ralph R., "Rapid Selector Patented," *Library J.*, 78, 310 (1953). Abstract of a patent (U. S. 2,594,358) for a Rapid Selector which scans 2000 subjects per second and copies the desired material as it scans. C.L.5:2. KL13

614. Shera, J. H., "Effect of machine methods on the organization of knowledge," *Am. Doc.*, 3, 15-20 (January 1952). JKO
615. Sherman, Jack, "Physical Data on Hydrocarbons", *Petroleum Refiner*, 32, No. 10, 145-49 (1953). A detailed description of the punched card system set-up for the physical data available from the American Petroleum Institute Project 44. C.L.6:1. J2 6 14
616. Sittner, M. L., "Labor Saving Devices and Techniques of the Future," *Medical Library Assoc. Bull.*, 41, 60-62 (January 1953). Reviews the application of machines to library operations. N11
617. Smith, A. Lee, "The Preparation and Use of (Edge-) Punch Cards Containing Printed Spectra," *Better Analysis*, 4, No. 1, 7, 10 (1953). Suggests tracing Spectra onto cards with pantograph, also manufacturing the punched cards from the photographic prints. DL14
618. Smith, William H., Russell A. Brant and Marion S. Klein, "An Application of Business Machine (Punched Card) Technique to Stratigraphic and Coal Resources Studies," Information Circular No. 18, Division of Geological Survey, State of Ohio, Columbus, 26 pp. iv (1956). J16
619. "Solving the Indexing Problem—Punched Cards Provide an Answer," *Chem. Age*, (London) 68, 965-66 (1953). C.L.5:4. DJ1
620. Someren Van, Ernest H., "Notes on a Punched Card Indexing System for Spectrochemical Communications," *Appl. Spectroscopy*, 7, 89-90 (1953). This system, in use for twelve years, is adapted to single row punched cards. C.L.5:4. D14
621. Soper, Alan K., "Some Observations on the Use of (Edge-) Punched Cards for a Personal Reference File," *Aslib Proc.*, 7, No. 4, 251-8 (1955). Superimposed random coding. DF1
622. Staats, Joan, "A Classified Bibliography of Inbred Strains of Mice," (on Edge-punched cards), *Science*, 119, 295-6 (1954). D1 12
623. Steidle, Walter, "Zur Dokumentation in der UV-Spektroskopie," *Z. Naturforsch.*, 10 b, 284-290 (May 1955). The article, in German, proposes a punched card indexing scheme, with samples, for ultraviolet spectroscopy. DO14
624. Stephens, Irlene Roemer (Celanese Corp., of America, Summit, N. J.), "History and Literature of Mechanized Data Handling." Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. N
625. Stevens, R. F. and J. F. Brady, "Application of Punched Cards to Chemical Process Control," *Chem. Eng. Progr.*, 50, 493-6 (1954). J2 7 8
626. Stewart, Charles B., Jr. (E. I. du Pont de Nemours & Co., Charleston, W. Va.), "A Study of (IBM) Punched Card Techniques Applied to the Indexing of Chemical Plant Engineering Drawings." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September, 1953. C.L.7:3. JO7
627. Strathacopoulos, A. D., "Noun Phrase Analysis (in machine translation) Using the IBM 650 Computer," *Trend. Eng. Univ. Wash.*, 9, No. 2, 20-22 (April 1957). C.L.9:3. K17
628. Strieby, I. M., "The Pharmaceutical Library of the Future," *Medical Library Assoc. Bull.*, 41, 399-404 (October 1953). N11 12
629. Taine, Seymour I., (National Library of Medicine, Washington, D.C.) "Mechanization in Methods of Publishing Lists and Indexes." Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. JLO1

630. Tapia, Elizabeth W. and James H. Gardner (National Research Corp., Cambridge, Mass.) "Coding of Several Types of Chemical Patent Abstracts for (Edge) Punch Card Use." Presented at the 128th national meeting of the American Chemical Society, Minneapolis, September 1955. C.L.7:3.
DF4 6 13
631. Taube, Mortimer, "Comments on Practicality of Coordinate Indexing," *College and Research Libraries*, 15, 419-420 (October 1954). An answer to G. E. Randall. (no. 586) A.D.6:1. GN
632. Taube, Mortimer, "Specificity in Subject Headings and Coordinate Indexing," *Library Trends*, 219-23 (October 1952). GO
633. Taube, Mortimer, and associates, "Studies in Coordinate Indexing," vols. II and III, Documentation Inc., Washington, D.C., 111 and 165 pp. respectively (1954 and 1956). A.D.7:4. GN
634. Thiers, Ralph E., "A Punch Card System for the Bibliography of Analytical Chemistry," *Virginia J. Sci.*, 2 (New Series), No. 1, 28-45 (1951). Includes mathematical analysis and discussion of coding possibilities of edge-cards. Gives a code for analytical chemistry. C.L.4:3. DFI 3
635. Thomas, Carl O., "A Jig for Preparing Edge Punched Cards," *J. Chem. Educ.*, 34, 241 (1957). D
636. Thomas, George R., "The Preparation of Punched Cards for Indexing Information," *J. Chem. Educ.*, 29, 406 (1952). A method is described for making edge punched cards using a drill press. C.L.4:4. D
637. Thompson, H. W., "The Documentation of Molecular Spectra," *J. Chem. Soc.*, 1956, 4501-09. A discussion and explanation of a new service on edged-punched cards for infrared and Raman spectra. Published in English by Butterworths Scientific Publications, London, and in German by Verlag Chemie, Weinheim. C.L.8:4. DP2 14
638. Thorne, R. G., "The Efficiency of Subject Catalogs and the Cost of Information Searches," *J. Document.*, 11 (3), 130-48 (1948). A method for quantitative evaluation of the efficiency of indexes. C.L.8:1. N1 8
639. Thurstone, L. L., "The Edge-Marking Method of Analyzing Data," Report No. 48 of the Psychometric Laboratory (Feb. 1948). B2
640. Tovey, Henry, (Bjorksten Research Laboratories) and Julian F. Smith (Takoma Park, Maryland) "Filing the Research Report in Industry and Government." Presented at the 124th national meeting of the American Chemical Society Chicago, September 1953. A punched card system is described. C.L.5:3. DJ15 16
641. "Transmitting Punch Card Data by 'Telephone' May have Wide Uses," *Petroleum Processing*, 9, 460 (1954). HJ
642. Tyler, A. W., Myers, W. L., and Kuipers, J. W., "The application of the Kodak Minicard system to problems of documentation," *Am. Doc.*, 6, 18-30 (January 1955). L.K. KL
643. "U. S. Department of Commerce," News release Oct. 22, 1954. Dr. Vannevar Bush is chairman of a newly appointed Advisory Committee on Application of Machines to Patent Office Operations. Suggestions are welcomed and should be sent to the secretary, Norman T. Ball, Room 2322, Department of Commerce. Washington 25, D.C. C.L.7:1. H13
644. do Valle, José Ribeiro, "Bibliographic Method for Using Punched Cards," Escola Paulista de Medicina, Sao Paulo, Brazil, 10 pp. (1955). D12
645. Van Atta, Fred F., "Indexing Infra Red Absorption Data on I.B.M. Cards," Circular, Am. Soc. for Testing Materials, Philadelphia (1953). A.S.T.M.

Committee E-13 is preparing to sell sets of cards made according to Kuentzel system. Included will be spectra issued by the American Petroleum Institute Research Project 44, by the National Research Council Committee on Spectral Absorption Data and by Samuel P. Sadtler and Son, Inc. C.L.6:2.

JP2 14

646. Vand, V., "Calculating Machines," *Nature*, 172, 371 (29 August 1953). Review of a book entitled "Les Machines à Penser" by Louis Couffignal (Paris: Les Editions de Minuit, 1952). It opens with definitions of a machine and of thought, followed by a description of adding machines and punched card machines. A.D.5:1. NO
647. Vanossican, W. W., "The Identification of Wood Species with the Aid of the Hollerith System," *Blumea*, 6, No. 1, 229-42 (1948). J2 12
648. Vickery, B. C., "Some Comments on Mechanical Selection," *Am. Doc.*, 2, 102-107 (April 1951). L.K. C.L.4:2. FNO
649. Vickery, B. C., "Developments in Subject Indexing," *J. Document.*, 11, 1-11 (March 1955). The author describes certain trends to illustrate the fact that the techniques of alphabetical indexing, classification and mechanical selection are developing along the same lines and in time will become more efficient tools in reference service. C.L.7:3. L.K. NO
650. Viennot, P., "Punched-card Machines; Informed Machines," *Bull. Soc. chim. France*, D-17-20 (1952). C.A.46:5904. JN
651. Wachtel, Irma S., "Indexing Nuclear Data on Punched Cards (Preliminary Edition)," Circular TID-469, U. S. Atomic Energy Commission, Technical Information Service, Oak Ridge 8 pp iv (1951). D5
652. Wachtel, I., "A Punched Card Index for Nuclear Data," *Am. Doc.*, 3, 56-57 (January 1952). Peek-a-boo system on IBM cards. L.K. C.L.4:3. D5
653. Wadington, John P., (Titanium Div., National Lead Co., South Amboy, N. J.) "Unit Concept Coordinate Indexing." Presented at the 131st national meeting of the American Chemical Society, Miami, April 1957. C.L.9:1 G
654. Wakelin, J. H., H. W. Lambert, and D. J. Montgomery, "A Fiber Length Tester with Automatic Punched-Card Registration," *Textile Research J.*, 26, 665-72 (1956). C.L.9:1. J7
655. Wall, Robert E., Jr., "Engineering Progress in Machine Translation," *Trend in Eng.*, 8, 9-15, 27-28 (July 1956). C.L.9:1. K17
656. Wallace, E. C., "Use of the ASM-SLA Metallurgical Literature Classification System by Industrial Metallurgists," *Spec. Libra.*, 47, 114-16 (1956). C.L.8:3. O7
657. Warheit, I. A., "Evaluation of Library Techniques for the Control of Research Materials," *Am. Doc.*, 7, 267-75 (1956). Stresses the need for having a thorough understanding of one's own needs before installing any of the various indexing systems which are available in order to make the correct choice of a system. C.L.9:1. N1 11
658. Way, Katharine, "Data-type Abstracts," *Phys. Today*, 10, 17-18 (1957). The Nuclear Data Group of the National Research Council has started a new type of informative abstracts which give experimental results on atomic nuclei and some information on experimental procedure, printed on cards. C.L.9:3. D5
659. Weaver, W., "The Patent Office Problem," *Am. Doc.*, 6, 129-33 (1955). Recommends what should be done over and beyond what the Advisory Committee on Application of Machines to Patent Office Operations recommended. C.L.7:4. H13

660. Weil, Barbara H., and E. A. Clapp, "Systems Presentation. The Adaptation of the ASM-SLA Metallurgical Literature Classification for Use in a Machine-Sorted Punched Card System," Metals Research Laboratories, Electro-Metallurgical Co., a Division of the Union Carbide & Carbon Corp., Niagara Falls, (Mimeographed), 29 pp. (1957). **JP1 7**
661. Weil, B. H., and Barbara Hildenbrand, "Information on Fuel and Lubricant Additives from Machine-Sorted (Remington Rand) Punched Cards" (Presented at "Symposium on Information Systems," Western Reserve Univ., Cleveland, Ohio, April 15-16, 1957). Mimeographed, 19 pp. (1957). Selection and correlation can be made on the basis of composition of the additive, use or function and the petroleum product. Cost figures are given. **J2 6 7 8**
662. Welt, I. D., (Cardiovascular Literature Project, Division of Medical Sciences, National Academy of Sciences, National Research Council, Washington, D.C.) "Perspective Aspects of Data-Handling Mechanization Deserving Research." Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3. **HN**
663. Whaley, Fred R. (Tonawanda Research Laboratory, Linde Air Products Co., Union Carbide and Carbon Corp., Tonawanda, N.Y.) "A Deep Index of Internal Technical Reports (on IBM cards)." Presented at the 130th national meeting of the American Chemical Society, Atlantic City, September 1956. C.L.8:3. **J15**
664. "What Size Spectra Do You Want to Record?," *The Perkin-Elmer Instrument News*, 4, No. 4, 6 (1953). The abscissa and ordinate scales of the Perkin-Elmer Model 21 infra red spectrophotometer are adjustable so spectra may be recorded on IBM or McBee punched cards or on other forms. **DJ2 14**
665. "Which Notation?," *Chem. Eng. News*, 33, 2838-43 (1955). Report of testing the Dyson, Gruber, Silk and Wiswesser notation systems for molecular structural formulas by volunteer teams. Discusses the Crane-Berry composite system. **HM**
666. Wildhack, W. A., J. Stern, and W. Smith, "Documentation in Instrumentation," *Am. Doc.*, 5, 223-237 (October 1954). "Peek-a-boo" system. L.K. **E1 14**
667. Williams, Theodore J., and Arthur Rose, "A Solution to the Problem of Storage and Availability of Chemical Literature," *J. Chem. Educ.*, 29, 146-147 (1952). A punched microcard is described which can be used in existing IBM sorting machines. C.L.4:2. C.A.46:5903. **JL1**
668. Wise, C. S., "Multiple Word Coding vs. Random Coding for the Rapid Selector," *Am. Doc.*, 3, 223-25 (1952). A reply to Calvin N. Mooers. (no. 531.) C.L.5:2. **FN**
669. Wiswesser, W. J. (Willson Products, Inc., Reading, Pa.), "Literature Sources of Mammalian Toxicity Data—with Special Emphasis on Tabulating Machinery Applications". Presented at the 124th national meeting of the American Chemical Society, Chicago, September 1953. C.L.5:3. **J1 12**
670. Wiswesser, William J., "Simplified Chemical Coding for Automatic Sorting and Printing Machinery," Willson Products, Inc., Reading, Pa., 1951. A method of coding chemical compounds. C.L.3:4. C.A.46:1860. **FM6**
671. Wood, G. Congdon, and Isaac D. Welt, "A Multi-Indexed Machine-Sorted, Punch Card System for Pesticide Metabolism Data," *J. Agr. Food Chem.*, 4, 886-88 (1956). **J2 12**
672. Woodbury, David O., "Let 'ERMA' Do It," Harcourt, Brace and Company, New York, 1956. Discusses automatic data handling machines. A.D.8:2. **KN11 17**

673. Wright, N., C. W. Young, and Ascher Opler, "Punched Card Identification of Mixtures by Combined Use of Infrared-Band Positions and No-Band Regions," Symposium on Molecular Structure and Spectroscopy, Ohio State University, Columbus, June 1951. Abstract in *Anal. Chem.*, **23**, 1050 (1951).
J2 14
674. Yngve, V. H., "The Technical Feasibility of Translating Languages by Machine," *Elec. Eng.*, **75**, 994-99 (1956).
K17
675. Young, A., "The Use of Punched Cards in Science and Administration," *Advances of Sci.*, **10**, 439-443 (March 1954). A.D.5:4.
N1 2 8 16
676. Zeman, Paul D., "Punched-card (Edge-punched) Catalog of Mass Spectra Useful in Qualitative Analysis," *Anal. Chem.*, **22**, 920-2 (1950). C.A.44:10488.
D2 3 14
677. Zimmerman, O. T., (Department of Chemical Engineering, University of New Hampshire, Durham, N. H.) "Cost of Installing and Operating Data-Handling Machinery (Related to Value)". Presented at the 132nd national meeting of the American Chemical Society, New York, September 1957. C.L.9:3.
N8

Joint Author Index for this Bibliography Only

(Reference is to Bibliography Serial Numbers)

- | | |
|--|---|
| Allison, Nancy J., 516 | Fullerton, Paul, 486, 608 |
| Andersen, Per, 282 | Gardner, James H., 630 |
| Anderson, E. E., 506 | Garfield, Eugene, 294 |
| Bannister, Lucy A., 387 | Gregory, Kenneth F., 588 |
| Beaufait, L. J., Jr., 506 | Hainer, R. M., 482 |
| Behm, Harold J., 354 | Hildenbrand, Barbara A., 425, 661 |
| Belzer, Jack, 601 | Howell, W. W., 517 |
| Berry, Madeline M., 405, 567, 568, 571, 572, | Hoxeng, R. B., 520 |
| Booth, K. H. V., 323 | Jeffrey, G. A., 429 |
| Brady, J. F., 625 | Jenkins, P., 308 |
| Brant, Russell A., 618 | Johnston, Herrick L., 601 |
| Brown, George W., 481 | Kalbe, Hans, 417 |
| Busbee, Ralph, 566 | Keller, Mark, 472 |
| Butterfield, Peter H., 473 | Kent, Allen 500, 567, 568, 569, 571, 572, 590 |
| Casey, Robert S., 573 | King, Gilbert W., 332 |
| Christen, H., 280 | Klein, Marion S., 618 |
| Clapp, E. A., 660 | Koller, H. R., 408, 493, 494 |
| Clark, Graham M., 460 | Kresze, G., 306 |
| Clews, C. J. B., 451 | Krueger, G. L., 286 |
| Cochran, W., 451 | Kuchitsu, Kozo, 543 |
| Conrad, C. C., 370 | Kuipers, J. W., 642 |
| Crandall, George S., 579 | Laing, K. M., 321 |
| Donnay, J. D. H., 374 | Lambert, H. W., 654 |
| Dubbs, D. S., 400 | Lanham, B. H., 296, 297 |
| Efron, Vera, 472, 476 | Leibowitz, Jacob, 296, 297, 407, 408, 492, 493, 494 |
| Fallon, F., 286 | Lewis, C. S., 380 |
| Fisher, Norman G., 387 | Luehrs, F. U., 571 |
| Ford, Eileen, 486 | |

- Lynn, G. E., 381
Maierson, Alvin T., 354
Marchington, B., 485
Maurer, E. H., 295
Mellon, M. G., 564
Monroe, R. A., 455
Montgomery, D. J., 654
Myers, W. L., 642
Newman, S. M., 285
Oneal, G., 333
Opler, Ascher, 673
Osmon, O., Jr., 414
Palopoli, F. P., 286
Pfeffer, H., 408, 494
Poplawska, A., 447
Priestley, William, Jr., 483
Prindle, Robert F., 548
Ridenour, Louis N., 332, 481
Rose, Arthur, 667
Ross, A. M., 457
Schomaker, Verner, 474
Schon, Harold, 417
Schumann, E. L., 286
Sherman, Jack, 360
Smith, W., 666
Stegemann, Herman, 417
Stern, J., 666
Stillman, J. W., 441
Takacs, L., 559
Taube, Mortimer, 330 (Part II).
Theriault, F. R., 600
Turbin, Kenneth, 376
Viervoll, H., 282
Vissek, W. J., 455
Wachtel, Irma S., 371
Way, Katherine, 371
Weil, Ben H., 425
Welt, Isaac D., 671
Wheeler, K. W., 286
White, W. B., 487
Williams, Theodore J., 597, 598
Young, C. W., 673

AUTHOR INDEX

(A separate author index is provided on pages 671-72 for the "Bibliography on Uses of Punched Cards" which forms Part V of this book).

- Abderhalden, E., 548
 Abegg, R., 548
 Adler, Frits H., 128
 Advisory Committee on Application of Machines to Patent Office Operations, 268, 617
 Advisory Committee on Configurational Nomenclature, 526
 Aldrous, J. G., 28
 Alembert, J. LeR. D', 614
 Allen, C. F. H., 314
 Allform Buro-organization GmbH, 127
 Alnord, Dorothy, 285
 American Medical Association, 241, 244
 American Petroleum Institute, Research Project, 44, 177
 American Society for Metals—Special Libraries Association Joint Committee, 100, 101, 431
 American Society for Testing Materials, 184, 209
 Anderson, E. E., joint author. See: Lukens, H. R., Jr., Anderson, E. E., and Beaufait, L. J., Jr., 305
 Andrews, D. D., 267, 404, 617
 Andrews, D. D., and Newman, S. M., 273
 Anglicus, Bartholomaeus, 613
 Anthony, R. N., 635
 Aristotle, 613
 Arnason, T. J., Cumming, E., and Spinks, J. W. T., 376
 Association of College and Reference Libraries, 290, 291
 Astanasoff, J. V., and Brandt, A. E., 436
 Baatz, Silmer H. and Maurer, Eugene H., 281
 Baehne, G. W., 302, 533
 Bagley, P. R., 617
 Bagley, P. R. and Perry, J. W., 617
 Bailer, J. C., Jr., Heumann, K. F., and Seiferle, E. J., 470
 Bailey, C. F., joint author. See: Casey, Robert S., Bailey, C. F., and Cox, Gerald J., 95, 445
 Bailey, C. F., joint author. See: Cox, Gerald J., Bailey, C. F., and Casey, Robert S., 7
 Bailey, C. F., joint author. See: Cox, Gerald J., Casey, Robert S., and Bailey, C. F., 21, 441
 Bailey, M. F., 262
 Bailey, M. F., Lanham, B. E., and Leibowitz, J., 264, 273, 508
 Baker, A. W., Wright, N., and Opler, A., 146, 206
 Baker, Catherine, 524
 Ball, N. T., 540
 Barnes, R. B., Gore, R. C., Liddle, U., and Williams, V. Z., 177
 Batten, W. E., 127
 Baud, P., joint author. See: Grignard, Victor and Baud, P., 548
 Bauman, Michael, 614
 Beaufait, L. J., Jr., joint author. See: Lukens, H. R., Jr., Anderson, E. E., and Beaufait, L. J., Jr., 305
 Beauvais, Vindent de, 614
 Becher, J. J., 615
 Begun, George M., 28
 Beilstein, F. K., 548
 Bernier, C. L. and Crane, E. J., 524
 Berolsheimer, D. D., joint author. See: West, C. J. and Berolsheimer, D. D., 548
 Berry, Madeline M. and Perry, J. W., 468
 Berry, Madeline M., joint author. See: Crane, E. M. and Berry, Madeline M., 468
 Berry, M. M., joint author. See: Perry, J. W., Kent, Allen and Berry, M. M., 249, 395, 397, 399, 406, 431, 541, 636
 Beukelman, T. E., 215
 Beyer, E., joint author. See: Pietsch, E. and Beyer, E., 616
 Bien, Saul M., 307
 Bishop, C., 325
 Biunno, Vincent P., 322
 Black-Schaffer, B. and Rosahn, P. D., 436
 Blair, E. H., 524
 Bloetjes, Mary K., 337, 338
 Boerhave, H., 615
 Boethius, A. M. S., 613
 Bonino, J. J. and Laing, K. M., 304
 Booth, H. S., joint author. See: Burkhard, C. A., Rochow, E. G., Booth, H. S., and Hartt, J., 93
 Borst, J. M., joint author. See: Cooper, F. S. Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., 357
 Borst, J. M., joint author. See: Cooper, F. S., Liberman, A. M., and Borst, J. M., 357
 Boyar, Se, 614
 Bradford, S. C., 616
 Brandt, A. E., 433, 436
 Brandt, A. E., joint author. See: Astanasoff, J. V. and Brandt, A. E., 436
 Brant, Russell A., joint author. See: Smith, William A., Brant, Russell A., and Klein, Margaret S., 321
 Brattain, R. R., 180
 Brauer, A. and D'Ans, J., 551
 Breger, I., 326
 Bridge Hand of the Month, Inc., 334
 Brooks, A. A., 525
 Brown, Betty M., joint author. See: Crandall, G. S. and Brown, Betty M., 564
 Brown, D. F., 554
 Brown, G. B., 281
 Brown, G. E., 525
 Buchen, W., 613
 Busa, Roberto, 361
 Bush, V., 10, 464
 Burkhard, C. A., 93
 Burkhard, C. A., Rochow, E. G., Booth, H. S., and Hartt, J., 93
 Butterfield, P. H., joint author. See: Jones, W. S. and Butterfield, P. H., 339
 Campbell, D. J., 28
 Canning, R. G., 635
 Cannon, W. A., Jr., 318
 Capella, Martianus, 613
 Carver, H. C., 434
 Casey, Robert S. and Perry, J. W., 127, 402, 442
 Casey, Robert S., Bailey, C. F., and Cox, Gerald J., 95, 445
 Casey, Robert S., joint author. See: Cox, Gerald J., Bailey, C. F., and Casey, Robert S., 7

- C. F., and Casey, Robert S., 7
- Casey, Robert S., joint author. See: Cox, Gerald J., Casey, Robert S., and Bailey, C. F., 21, 441
- Casey, Robert S., joint author. See: Hill, Norman C., Casey, Robert S., and Perry, James W., 6
- Cassels, W. H., joint author. See: Conroy, W. Allen, Cassels, W. H., and Stodsky, Bernard, 174
- Chantimpre, Thomas de, 614
- Chappell, Mrs. James F., 321
- Chemical Codification Panel, National Research Council, 470
- Cherry, C., 617
- Chronic, John, 320
- Clark, C., 183
- Clark, J. M., 613
- Clarke, A. L., 525
- Clarke, S. H., 433
- Clavier, Henri, 540
- Clem, M. A., joint author. See: Homeyer, P. G., Clem, M. A., and Federer, W. T., 436
- Cline, Carl A., 498
- Cobb, Charles K., Jr., 322
- Cochran, S. W., 565
- Colinese, P. E., 101
- Commission de Nomenclature de Chimie Biologique, 526
- Commission de Nomenclature de Chimie Inorganique, 526
- Commission de Nomenclature de Chimie Organique, 526
- Committee on the Reform of the Nomenclature of Organic Chemistry, 525
- Commission des Symboles et de Terminologie Physico-chimique, 527
- Committee on Biochemical Nomenclature, National Research Council, 526
- Committee for the Reform of Inorganic Chemical Nomenclature, 525
- Committee on Spectral Absorption Data, National Research Council, 177
- Conroy, W. Allen, Cassels, W. H., and Stodsky, Bernard, 174
- Consolidated Electrodynamics Corp., 223
- Cook, C. M., 373
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., 357
- Cooper, F. S., Liberman, A. M., and Borst, J. M., 357
- Cordonnier, G., 127
- Couffignal, L., 617
- Cox, Gerald J., Bailey, C. F., and Casey, Robert S., 7
- Cox, Gerald J., Casey, Robert S., and Bailey, C. F., 21, 441
- Cox, Gerald J., joint author. See: Casey, Robert S., Bailey, C. F. and Cox, Gerald J., 95, 445
- Crandall, G. S., and Brown, Betty M., 564
- Crane, E. M., 616
- Crane, E. J., Patterson, A. M., and Marr, Eleanor B., 525
- Crane, E. J., joint author. See: Bernier, C. L. and Crane, E. J., 524
- Crane, E. M. and Berry, Madeline M., 468
- Cullman, Ralph E., 29
- Cumming, E., joint author. See: Arnason, T. J., Cumming, E., and Spinks, J. W. T., 376
- D'Ans, J. and Brauer, A., 551
- Davis, L. R., 334
- Davison, W. H. T., 489
- Davison, W. H. T., joint author. See: Gordon, M., Kendall, C. E., and Davison, W. H. T., 487
- Dangl, J. R., joint author. See: Randall, H. M., Fowler, R. G., Fuson, N., and Dangl, J. R., 177
- De Gorter, B., 127, 561
- Delattre, P. C., joint author. See: Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gertman, L. J., 357
- De Mamantoff, N., 127
- Demer, L. J., 339
- Dennis, Sally F., 405
- Detectri, 127
- De War, D., 319
- Dewey, John, 536
- Dewey, Melvil, 240
- Diderot, Denis, 614
- Diebold, Inc., 289
- Dobriner, K., Katzenellenbogen, E. R., and Jones, R. N., 177
- Documentation, Inc., 129
- Donney, J. D. H., 126
- Dooher, M. J., 636
- Doss, Milburn Price, 558
- Drillick, J. H., 127
- Duer, Margaret M., and Lewis, Clark S., 280
- Dunn, E. E. and Lynn, G. E., 305
- Dunn, H. L., joint author. See: Tovell, R. M. and Dunn, H. L., 174
- Dunston, et al., 549
- Dyson, G. M., 476
- Eastman Kodak Co., 312
- Eckart, E., 4
- Eckert, W. J., 435
- Eckert, W. J., and Haupt, R. F., 566
- Eckert, W. J., and Jones, R., 618
- Edelman, David L., 102
- Edler and Krische, 127
- Edwards, T. I., 436
- Engineering Research Associates, Inc., 567
- Engler, C. and Hoefler, E. v. H., 549
- Erman, A., 612
- Eudemos, 613
- Ezekiel, M., 433
- Fairthorne, R. A., 431, 540
- Faust, O., 551
- Federer, W. T., joint author. See: Homeyer, P. G., Clem, M. A., and Federer, W. T., 436
- Fehling, H., 548
- Fisher, R. A. and Yates, P., 236
- Flagg, C. R., joint author. See: Kent, Allen and Flagg, C. R., 249
- Fletcher, W. I., 525
- Foster, L. S., 101
- Fowler, R. G., joint author. See: Randall, H. M., Fowler, R. G., Fuson, N., and Dangl, J. R., 177
- Frear, Donald E. H., 7, 435, 468
- Frear, D. E. H. and Seiferle, E. J., 469
- Frear, D. E. H., and Seiferle, E. J., and King, H. L., 468
- Frear, D. E. H., joint author. See: Morgan, J. A. and Frear, D. E. H., 470
- Frevel, L. K., joint author. See: Hanawalt, J. D., Rinn, H. W., and Frevel, L. K., 183
- Friedlander, P., 551
- Friend, J. A. N., 548
- Fritze, E., 613
- Frome, Julius and Leibowitz, Jacob, 404
- Fulcher, G. S., 525

- Fuson, N., joint author. See: Randall, H. M., Fowler, R. G., Fuson, N., and Dangel, J. R., 177
- Gagarin, R., 618
- Geer, Harriet, joint author. See: Kent, Allen and Geer, Harriet, 11
- Geer, Helen T., 298, 302
- Geisler, A. H., 100
- Geisler, A. H., joint author. See: Guy, A. G., and Geisler, A. H., 100, 460
- Gent, E. W., et al., 127
- Gerardin, A., 125
- Gerstman, L. J., joint author. See: Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., 357
- Gey, K. F., Kalbe, H., Schon, H., and Stegemann, H., 339
- Gillen, H. W., 337
- Glaser, C., 615
- Gmelin, Johann Friedrich, 615
- Gmelin, L., 548, 551, 615
- Gohkle, R. S., joint author. See: McLafferty, F. W. and Gohkle, R. S., 226
- Goldman, S., 617
- Good, Carter V. and Scates, Douglas E., 540
- Gordon, A. H., 319
- Gordon, M., Kendall, C. E., and Davison, W. H. T., 487
- Gore, R. C., joint author. See: Barnes, R. B., Gore, R. C., Liddle, U., and Williams, V. Z., 177
- Graham, M. H., Hildenbrand, B. S., and Weil, Ben H., 333
- Gray, C. J., 126
- Gregg, J. R., 540
- Grignard, Victor and Baud, P., 548
- Gruber, W., 479
- Guirand, Pierre, 358
- Gull, C. D., 302, 561, 564
- Gull, C. D., joint author. See: Kreithin, A., Gull, C. D., and Miller, E. E., 618
- Guy, A. G., and Geisler, A. H., 100, 460
- Hale, A. H. and Stillman, J. W., 313
- Hampton, R. R., 180
- Hanawalt, J. D., Rinn, H. W., and Frevel, L. K., 183
- Hantke, G., joint author. See: Pietsch, E., Hantke, G., and Mulert, G., 612
- Harden, Florence, 552
- Harff, H., 616
- Harms, F., joint author. See: Wien, W., and Harms, F., 548
- Hartt, J., joint author. See: Burkhard, C. A., Rochow, E. G., Booth, H. S., and Hartt, J., 93
- Haskins and Sells, 636
- Hass, H. B., 616
- Haupt, R. F., joint author. See: Eckert, W. J. and Haupt, R. F., 566
- Heffter, A., 548
- Helmont, J. B. van, 615
- Hempel, E. H., 553
- Hensel, Evelyn, et al., 558
- Herberts, K., 612
- Hetherington, E., 525
- Heumann, K. F., 635
- Heumann, K. F., joint author. See: Bailar, J. C., Jr., Heumann, K. F., and Seiferle, E. J., 470
- Hildenbrand, B. S., joint author. See: Graham, M. H., Hildenbrand, B. S., and Weil, B. H., 333
- Hill, Albert G., joint author. See: Ridenour, Louis, N., Shaw, Ralph R., and Hill, Albert G., 398
- Hill, Norman C., Casey, Robert S., and Perry, James W., 6
- Hill, Thomas T., 312, 333
- Hocker, Margaret L., 299
- Hoefer, E. v. H., joint author. See: Engler, C. and Hoefer, E. v. H., 549
- Hoffman, E. J., 333
- Hohenhoff, E., 551
- Holden, W. H. T., 431
- Holmstrom, E., 618
- Homeyer, P. G., Clem, M. A., and Federer, W. T., 436
- Hood, M. and Lyle, G. R., 297
- Hoesmann, H., 613
- Houben, 551
- Houchin, Thomas D., Puhvel, Jaan, and Watkins, Calvert W., 358
- Hoxeng, R. B., joint author. See: Mathay, W. L. and Hoxeng, R. B., 329
- Humboldt, Alexander von, 614
- Hyslop, Marjorie R., 249, 398, 613
- International Business Machines Corp., 280, 302, 563, 564, 565
- Isbell, A. F., 458
- Isidore of Seville, 613
- Italian Association of Metallurgy, 101
- Jeffers, H. M., 321
- Jevons, W. S., 533, 541
- Jones, R., joint author. See: Eckert, W. J. and Jones, R., 618
- Jones, R. N., joint author. See: Dobriner, K., Katzenellenbogen, E. R., and Jones, R. N., 177
- Jones, W. S. and Butterfield, P. H., 339
- Jungk, R., 612
- Jungken, J. H., 615
- Kalbe, H., joint author. See: Gey, K. F., Kalbe, H., Schon, H., and Stegemann, H., 339
- Katsenellenbogen, E. R., joint author. See: Dobriner, K., Katzenellenbogen, E. R., and Jones, R. N., 177
- Kaufman, A., 614
- Keckley, C. P., 440
- Keller, Alton H., 284
- Kendall, C. E., 216
- Kendall, C. E., joint author. See: Gordon, M., Kendall, C. E., and Davison, W. H. T., 487
- Kent, Allen, 9, 11
- Kent, Allen and Flagg, C. R., 249
- Kent, Allen and Geer, Harriet, 11
- Kent, Allen and Perry, J. W., 249, 250
- Kent, Allen, joint author. See: Peakes, G. L., Kent, A., and Perry, J. W., 339
- Kent, Allen, joint author. See: Perry, J. W. and Kent, Allen, 249, 250, 252, 253, 395, 397, 399, 405, 406
- Kent, Allen, joint author. See: Perry, J. W., Kent, Allen, and Berry, M. M., 249, 395, 397, 399, 406, 431, 541, 636
- Kent, Allen, joint author. See: Shera, J. H., Kent, Allen, and Perry, J. W., 249, 339, 396, 398, 420, 613, 616, 617
- Kilgour, F. G., 302
- King, Gilbert W., 10, 618
- King, H. L., joint author. See: Frear, D. E. H., Seiferle, E. J., and King, H. L., 468
- Kirk, R. E., and Othmer, D. F., 549

- Kirsch, R. A., joint author. See: Ray, L. C., and Kirsch, R. A., 275
- Kirschner, S., 326
- Klein, Margaret S., joint author. See: Smith, William A., Brant, Russell A., and Klein, Margaret S., 321
- Klingman, H. F., 636
- Koller, H. R., joint author. See: Lanham, B. E., Leibowitz, J., and Koller, H. R., 265, 489
- Koller, H. R., joint author. See: Lanham, B. E., Leibowitz, J., Koller, H. R., and Pfeffer, H., 278
- Koller, H. R., joint author. See: Pfeffer, H., Koller, H. R., and Marden, E., 278
- Konigschlafer, Peter, 614
- Kopp, H., 615
- Kounts, R. R., 339
- Kraitichich, Maurice, 125, 126
- Kraut, K., 615
- Kreithin, A., Gull, C. D., and Miller, E. E., 618
- Krull, A. R., 331
- Kubach, I., 618
- Kuentzel, L. E., 183, 211, 216
- Kuhn, H., 612
- Kuipers, J. W., joint author. See: Tyler, A. W., Myers, W. L., and Kuipers, J. W., 396, 617
- Kyle, Barbara, 541
- Lafontaine, H., and Otlet, P., 613
- Laing, K. M., joint author. See: Bonino, J. J., and Laing, K. M., 304
- Lanham, B. E., 264, 267
- Lanham, B. E., Leibowitz, J., and Koller, H. R., 265, 489
- Lanham, B. E., Leibowitz, J., Koller, H. R., and Pfeffer, H., 278
- Lanham, B. E., joint author. See: Bailey, M. F., Lanham, B. E., and Leibowitz, J., 264, 273, 508
- Lavoisier, A. L., 615
- Leary, J. C., joint author. See: MacKinnon, F. B., Leary, J. C., and Levinson, D., Jr., 322
- Lefebvre, N., 615
- Leibniz, G. W., von, 614
- Leibowitz, J., joint author. See: Bailey, M. F., Lanham, B. E., and Leibowitz, J., 264, 273, 508
- Leibowitz, Jacob, joint author. See: Frome, Julius and Leibowitz, Jacob, 404
- Leibowitz, J., joint author. See: Lanham, B. E., Leibowitz, J., and Koller, H. R., 265, 489
- Leibowitz, J., joint author. See: Lanham, B. E., Leibowitz, J., Koller, H. R., and Pfeffer, H., 278
- Lemery, N., 615
- Leniham, J. M. A., 339
- Lester, A. M., 28
- Levine, Norman D., 308
- Levinson, D., Jr., joint author. See: MacKinnon, F. B., Leary, J. C., and Levinson, D., Jr., 322
- Lewis, C. M., 616
- Lewis, Chester M. and Offenhauser, William H., 558
- Lewis, Clark S., joint author. See: Duer, Margaret M., and Lewis, Clark S., 280
- Libavius, A., 615
- Liberman, A. M., 357
- Liberman, A. M., joint author. See: Cooper, F. S. Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., 357
- Liberman, A. M., joint author. See: Cooper, F. S., Liberman, A. M., and Borst, J. M., 357
- Liddle, U., joint author. See: Barnes, R. B., Gore, R. C., Liddle, U., and Williams, V. Z., 177
- Lieber, Henri, 127
- Lindenmeyer, H. F., 265
- List, K., 615
- Lockhart, E. E., 336
- Los Angeles Police Dept., 336
- Low, Ward C., 318
- Lowe, Ruth K., 302
- Luhn, H. P., 394, 405, 491, 492, 499, 506
- Lukens, H. R., Jr., Anderson, E. E., and Beaufait, L. J., Jr., 305
- Lutz, A. W., 565
- LuValle, James E., 311
- Lyle, G. R., joint author. See: Hood, M., and Lyle, G. R., 297
- Lynn, G. E., joint author. See: Dunn, E. E., and Lynn, G. E., 305
- MacKinnon, F. B., Leary, J. C., and Levinson, D., Jr., 322
- Macquer, D. J., 615
- Magalhães, Hulda, 308
- Marden, E., joint author. See: Pfeffer, H., Koller, H. R., and Marden, E., 278
- Mark, H., and Proskauer, E. S., 551
- Marr, Eleanor B., joint author. See: Crane, E. J., Patterson, A. M., and Marr, Eleanor B., 525
- Martin, Nella Jane, 240
- Marting, E., 636
- Mathay, W. L., and Hoxeng, R. B., 329
- Matthews, F. W., 209
- Maurer, Eugene H., joint author. See: Baatz, Silmer H. and Maurer, Eugene H., 281
- Maxfield, D. K., 302
- McCrea, J. M., 226
- McCrone, W. C., 211
- McCulley, William R., 568
- McGraw, Howard F., 279
- McJoynt, Joan R., 335
- McLafferty, F. W., and Gohlke, R. S., 226
- Megenberg, Konrad von, 614
- Mellor, J. W., 549
- Melton, Jessica, 250
- Melton, Jessica and Perry, J. W., 395, 397
- Meyer, R. J., 616
- Miller, Eugene, 129
- Miller, E. E., joint author. See: Kreithin, A., Gull, C. D., and Miller, E. E., 618
- Milne, Lorus J., and Margery, 311
- Milne, Margery, joint author. See: Milne, Lorus J. and Margery, 311
- Monkevich, Edward, 288
- Mooers, Calvin N., 233, 344, 457, 490, 636
- Mooers, Calvin N. and Charlotte D., 341
- Mooers, Charlotte D., joint author. See: Mooers, Calvin N. and Charlotte D., 341
- Morgan, J. A., and Frear, D. E. H., 470
- Mulert, G., joint author. See: Pietsch, E., Hantke, G., and Mulert, G., 612
- Mulert, G., joint author. See: Pietsch, E. and Mulert, G., 618
- Myers, O., 126
- Myers, W. L., joint author. See: Tyler, A. W., Myers, W. L., and Kuipers, J. W., 396, 617
- Naimark, G. M., 313
- Naimark, G. M. and Prindle, R. F., 313

- National Association of Corrosion Engineers, 339
 National Science Foundation, 335, 339
 Newman, S. M., 265
 Newman, S. M., joint author. See: Andrews, D. D. and Newman, S. M., 273
 Nicholas, J. B., 515
 Nichols, J. B., 525
 Norton, T. R., and Opler, A., 480
 Norton, T. R., joint author. See: Opler, A. and Norton, T. R., 480, 482, 636
 Noeworthy, Michael, 174
- Offenhauser, William H., joint author. See: Lewis, Chester M., and Offenhauser, William H., 558
 Ohrmund, Margaret, 306
 Opler, A. and Norton, T. R., 480, 482, 636
 Opler, A., joint author. See: Baker, A. W., Wright, N., and Opler, A., 146, 206
 Opler, A., joint author. See: Norton, T. R., and Opler, A., 480
 Oppenheimer, K., 548
 Orr, C. H., 324
 Othmer, D. F., joint author. See: Kirk, R. E., and Othmer, D. F., 549
 Otis, M. V., 206
 Otlet, P., 612
 Otlet, P., joint author. See: Lafontaine, H. and Otlet, P., 613
- Paracelsus, 615
 Parker, Ralph H., 279, 302
 Pascal, P., 458
 Passer, Moses, 560
 Paterson, D. D., 433
 Patterson, A. M., joint author. See: Crane, E. J., Patterson, A. M., and Marr, Eleanor B., 525
 Patton, A. R., 334
 Peakes, G. L., 339, 566
 Peakes, G. L., Kent, A., and Perry, J. W., 339
 Pender, John W., 174
 Perkins, F. B., 525
 Perry, J. W., 402, 405
 Perry, J. W., and Kent, Allen, 249, 250, 252, 253, 395, 397, 399, 405, 406
 Perry, J. W., Kent, Allen, and Berry, M. M., 249, 395, 397, 399, 406, 431, 541, 636
 Perry, J. W., joint author. See: Bagley, P. R. and Perry, J. W., 617
 Perry, J. W., joint author. See: Berry, Madeline M., and Perry, J. W., 468
 Perry, J. W., joint author. See: Casey, Robert S., and Perry, J. W., 127, 402, 442
 Perry, James W., joint author. See: Hill, Norman C., Casey, Robert S., and Perry, James W., 6
 Perry, J. W., joint author. See: Kent, Allen and Perry, J. W., 249, 250
 Perry, J. W., joint author. See: Melton, Jessica and Perry, J. W., 395, 397
 Perry, J. W., joint author. See: Peakes, G. L., Kent, A., and Perry, J. W., 339
 Perry, J. W., joint author. See: Shera, J. H., Kent, Allen, and Perry, J. W., 249, 339, 396, 398, 420, 613, 616, 617
 Perry, J. W., joint author. See: Wise, Carl S., and Perry, J. W., 398, 464, 567
 Pfeffer, H., Koller, H. R., and Marden, E., 278
 Pfeffer, H., joint author. See: Lanham, B. E., Leibowitz, J., Koller, H. R., and Pfeffer, H., 278
 Pfucke, M., 616
 Pietsch, E., 612, 613, 614, 616, 617, 618
 Pietsch, E., and Beyer, E., 616
 Pietsch, E., Hantke, G., and Mulert, G., 612
 Pietsch, E., and Mulert, G., 618
 Pinkerton, J. M. M., 567
 Piper, R. F., and Ward, P. W., 541
 Plinius, Caius Secundus, 613
 Pollard, A. W., 525
 Prindle, R. F., joint author. See: Naimark, G. M., and Prindle, R. F., 313
 Proskauer, E. S., joint author. See: Mark, H., and Proskauer, E. S., 551
 Puhvel, Jaan, joint author. See: Houchin, Thomas D., Puhvel, Joan, and Watkins, Calvert W., 358
- Quigley, Margery, 280
- Radioactive Products, Inc., 304
 Randall, H. M., Fowler, R. G., Fuson, N., and Dangel, J. R., 177
 Ray, L. C., and Kirsch, R. A., 275
- Reichardt, G., 616
 Reumuth, H., 28, 339
 Rhynders, R. W., 315
 Richardson, E. C., 541
 Richter, F., 616
 Ridenour, Louis N., Shaw, Ralph R., and Hill, Albert G., 398
 Rift, Leo R., 302
 Rinn, H. W., joint author. See: Hanawalt, J. D., Rinn, H. W., and Frevel, L. K., 183
 Robinson, O. H., 525
 Robinson, R. M., 128
 Rochow, E. G., joint author. See: Burkhard, C. A., Lochow, E. G., Booth H. S., and Hartt, J., 93
 Rosa, M. C., 264, 535
 Rosahn, P. D., joint author. See: Black-Schaffer, B. and Rosahn, P. D., 436
 Rose, A., joint author. See: Williams, T. J. and Rose, A., 323
 Rossini, F. D., 482
 Rovenstine, E. A., 174
 Rowe, F. M., 551
 Ruston, W. R., 618
- Sadtler, Samuel P., and Son, Inc., 177
 Samain, J., 617
 Savitsky, Abraham, 146
 Scates, Douglas E., joint author. See: Good, Carter V. and Scates, Douglas E., 540
 Scheele, M., 613
 Schlink, F. J., 28
 Schon, H., joint author. See: Gey, K. F., Kalbe, H., Schon, H., and Stegemann, H., 339
 Schultz, G., 551
 Schulz, Hugo, 614
 Schurmeyer, Walther, 127
 Schwabe, C. W., and Davis, L. R., 339
 Scott, E., 617
 Seely, J. S., 129
 Seiferle, E. J., joint author. See: Bailar, J. C., Jr., Heumann, K. F., and Seiferle, E. J., 470
 Seiferle, E. J., joint author. See: Frear, D. E. H., and Seiferle, E. J., 469
 Seiferle, E. J., joint author. See: Frear, D. E. H., Seiferle, E. J., and King, H. L., 468
 Shaw, Ralph R., 398, 617
 Shaw, Ralph R., joint author. See: Ridenour, Louis N., Shaw, Ralph R., and Hill, Albert G., 398

- Shepardson, G. D., 525
 Shera, J. H., Kent, Allen, and Perry, J. W., 249, 339, 396, 398, 420, 613, 616, 617
 Sherman, J., 331
 Siegel, W., 549
 Smith, J., joint author. See: Wildhack, W. A., Stern, J., and Smith, J., 125
 Smith, S. W., 302
 Smith, William A., Brant, Russell A., and Klein, Marian S., 321
 Soper, H., 126
 Soule, G., 373
 Sparrow, W. K., 4
 Sperry-Rand Corp., 128
 Spinks, J. W. T., joint author. See: Arnason, T. J., Cumming, E., and Spinks, J. W. T., 376
 Staats, H. N., 339
 Staats, J., 306
 Stahl, G. E., 615
 Stegemann, H., joint author. See: Gey, K. F., Kalbe, H., Schon, H., and Stegemann, H., 339
 Steier, H. P., 618
 Stern, J., joint author. See: Wildback, W. A., Stern, J., and Smith, J., 125
 Stillman, J. W., joint author. See: Hale, A. H., and Stillman, J. W., 313
 Stodsky, Bernard, joint author. See: Conroy, W. Allen, Casals, W. H., and Stodsky, Bernard, 174
 Stokes, Katharine M., 299
 Stroupe, James D., 180
 Superintendent of Documents, 262, 263
 Tasman, P., 359
 Taube, Mortimer, 560, 561, 618
 Taube, Mortimer and Wachtel, Irma S., 431
 Taube, M., et al., 636
 Taylor, F. Lowell, 284
 Taylor, H., 125
 Theophrastus, 613
 Thomas, Carl O., 28
 Thomas, George R., 28
 Thompson, H. W., 183
 Thomson, W. S., 525
 Thorpe, J. F., 549
 Thurstone, L. L., 28
 Todd, A. R., 616
 Tovell, R. M., and Dunn, H. L., 174
 Tachirch, 549
 Tyler, A. W., Myers, W. L., and Kuipers, J. W., 396, 617
 Tyler, C., 553
 Ullmann, F., 549
 Underwood Corp., 129, 289
 U. S. Navy, Bureau of Aeronautics, 338
 U. S. Patent Office, 263, 264, 267
 Universal Oil Products Corp., 551
 Wachtel, Irma S., 303, 304
 Wachtel, Irma S., joint author. See: Taube, Mortimer and Wachtel, Irma S., 431
 Wallace, E. C., 102
 Ward, P. W., joint author. See: Piper, R. F., and Ward, P. W., 541
 Warren, I., 525
 Wassell Organization, Inc., 137
 Wassenberg, Alvina, 102
 Watkins, Calvert W., joint author. See: Houchin, Thomas D., Puhvel, Jaan, and Watkins, Calvert W., 358
 Way, K., 339
 Weil, B. H., joint author. See: Graham, M. H., Hildenbrand, B. S., and Weil, B. H., 333
 Wellensiek, S. J., 384
 Welt, I. D., joint author. See: Wood, G. C., and Welt, I. D., 311
 West, C. J., and Berolzheimer, D. D., 548
 Wheatley, H. B., 525
 Wheeler, M. T., 525
 Wiegble, J. C., 615
 Wien, W., and Harms, F., 548
 Wiener, N., 617
 Wildback, W. A., Stern, J., and Smith, J., 125
 Williams, T. J., and Rose, A., 323
 Williams, V. Z., joint author. See: Barnes, R. B., Gore, R. C., Liddle, U., and Williams, V. Z., 177
 Wilson, H. W., 525
 Wilson, H. W., Co., 302
 Wise, Carl S., 233, 402, 441, 442, 464
 Wise, Carl S., and Perry, J. W., 398, 464, 567
 Wiswesser, W. J., 482
 Wood, G. C., and Welt, I. D., 311
 Worden, E. C., 551
 Wright, N., 180
 Wright, N., joint author. See: Baker, A. W., Wright, N., and Opler, A., 146, 206
 Yates, P., joint author. See: Fisher, R. A., and Yates, P., 236
 Zemany, P. D., 225

SUBJECT INDEX

- Abstract periodicals
 - and card files, 93-94, 99, 112, 114-115, 379-386
 - as guide to original literature, 543-544
 - historical notes, 579-580
 - indexing problems, 387, 510-524, 546
 - limitations of scope, 552
 - in plant breeding and genetics, 374-387
 - statistics on growth and expansion, 580-581, 591-592
- Abstracts, encoded, as basis for searching, 249-252
- Accounting. See also *Computation (operations and problems)*
 - edge-punched card equipment for, 30-46, 49-52
 - IBM equipment for, 54-64
 - Remington-Rand equipment for, 64-74
- Accounting machine, IBM system, 59-60
- Accounting machine cards, used as hand-sorted cards, 128-129, 226, 303
- Acquisition work in libraries. See *Ordering routines in libraries*
- Addressograph copying, and edge-punched or hand-sorted cards, 558
- Adhesives. See also *Edge-punched cards; Mounting of materials.*
 - for mounting material on cards, 27
- Alignment block, 13-17, 32-33, 46-47
- Alphabetical coding. See also *Punching schemes*
 - of author names, 95-96, 110-111, 497-498
 - based on letter use frequencies, 497-499
 - for data processing equipment, 625
 - for edge-punched cards, 20
 - with IBM cards, 6, 54-55, 494-504
 - mathematical analysis, 440-443
 - with Remington-Rand cards, 6, 64-65
 - and superimposed word coding, 447-464, 494-504
- Alpha-matrix system, cards and equipment, 85-88
- Alphanumeric coding for use in computers, 625
- Aluminum and aluminum alloys. See also *Metallurgy; Metals and alloys*
 - coded on edge-punched cards, 112, 114-117
- American Society for Metals-Special Libraries Association classification system. See *ASM-SLA Metallurgical Literature Classification*
- American Society for Testing Materials, punched card systems for spectral data, 183-207, 209-215, 216-223, 225-231
- Analysis by spectral methods, aided by punched cards, 175-231
- Analysis of information. See *Code development (and use); Subject analysis*
- Analytical chemistry data, on Keysort cards, 312-314
- Ancillary equipment, list of manufacturers of, 90
- Anesthesia records
 - advantages of punched-card file, 161-162
 - codes for punched card system, 163-173
 - on edge-punched cards, 161-174
 - use of punched cards for statistical analyses, 161, 173
- Antibiotics literature file, with punched card system, 306
- ASM-SLA Metallurgical Literature Classification
 - experimental searching system, 248-260
 - general discussion, 102-105
 - and hand-sorted punched cards, 102, 105-112
 - history, 100-102, 248-250
 - provision for expansion, 111-112
 - revised edition, 118-124
 - workbook, 112
- Astronomical data, with punched card systems, 321
- Author names. See also *Alphabetical Coding*
 - alphabetical coding of, 95-96, 110-111, 324, 497-598
 - coding based on frequency of occurrence of letters, 110-111
 - coded by consonant code, 497-498
 - coded on edge-punched cards, 95-96, 110-111, 324, 354-355

- in peek-a-boo system, 145-146
- scatter coding of, 354-355
- Automatic typing and IBM cards, 562, 566
- Autopsies, correlation of, 436
- Berry-Crane composite code for organic compounds**, 485-487
- Bibliography preparation**
 - with edge-punched cards, 19, 25, 98-99, 116
 - introductory discussion, 6-7, 10-11
 - in microprint with "Microcite" copying system, 147-149
 - with Powers-Samas cards, 384-386
 - from punched or hand-sorted cards, and reproduction techniques, 561-562, 564, 566
- Binary code as definition of information content of punched cards**, 423-424
- Binary numbers**
 - relationship to alphanumerical coding, 625
 - relationship to numerical coding, 441-445
 - used in computer codes, 624-625
- Bindery records in libraries**, use of edge-punched card systems, 285
- Biological data files**, with punched card systems, 232-247, 305, 311
- Blueprints of engineering drawings**, with Filmsort system, 338
- Boiling point**, coded on punched cards, 178, 205, 224
- Book budget accounting in libraries**, with IBM system, 281
- Book cards**. See *Circulation control in libraries*
- Books**
 - classification of, 529-530, 544
 - guides to, 544
 - lists and literature searching of, 548-549
 - reviews and literature searching of, 549
- Bridge hands**, filed with punched card system, 333-334
- Calculating punch**, IBM system, 61-62
- Calculating punch**, Remington-Rand system, 70-71
- Call cards**. See *Circulation control in libraries*
- Card design**. See *Punched cards, design and layout*
- Card files**, conventional
 - inadequacies of title indexing, 152
 - in libraries, limitations of, 544
- Card-programmed electronic calculator**, IBM system, 62
- Card savers**, 31-32, 38, 49
- Cardatype machines (IBM)**, 361, 563, 565-566
- Catalog lists**, production of, with punched cards, 285-286
- Cataloging in libraries**, use of machine-sorted cards, 285-287
- Catalyst data**, with punched card system, 333
- Charging systems**. See *Circulation control in libraries*
- Chemical Abstracts**. See *Abstract periodicals*
- Chemical compounds**
 - Chodosch code, 490
 - and classifying, indexing, and coding, 266-267, 465-468
 - coded on IBM cards, 187-202, 214, 227-228, 229-231
 - coded on Remington-Rand cards, 332
 - coded for spectral data punched card systems, 178-183, 187-205, 210, 213-214, 224-225, 228-231
 - coding schemes for, general review, 465-491
 - empirical formulas coded on IBM cards, 227-228
 - Frear code, 468-470
 - Gordon-Kendall-Davison code, 487-489
 - inorganic, code for, 597-602
 - names coded on IBM cards, 227-228, 229-321
- National Research Council Code**, 470-476
- Nodal index code**, 491
- organic**
 - Berry-Crane composite code, 485-487
 - coded on IBM cards, 187-202, 214
 - coded on Keysort cards, 180-183, 313-314
 - Dyson code, 476-479
 - Gruber code, 479

- Opler and Norton (Dow) code, 480-482
random number code, 243-244
Wiswesser code, 482-485
organosilicon, code for, 94-95
properties coded, for punched card systems, 178, 185-187, 205, 212-214, 217-219, 221-225
radicals coded on IBM cards, 204-205, 213
relationship between indexing, classifying and coding, 465, 468
U. S. Patent Office codes, 270, 274, 489-490
Zatopleg code, 490-491
- Chemical elements
coded on IBM cards, 187-188, 202-203, 213, 597-599, 601-602
coded on edge-punched cards, 108-110, 210-211, 325-326
- Chemical literature files. See also specific subjects, as *Chemical compounds*; *Laboratory records*; *Petroleum data*; etc.
with punched card systems, general review, 6, 7, 323-329
- Chodosch code for chemical compounds, 490
- Chronological codes. See *Date, coding of*
Circulation control in libraries
use of edge-punched cards, 290-291, 296-299
general discussion, 287-288
use of machine-sorted cards, 288-297
- Classification of cards. See *Filing cards in groups*
- Classification systems. See also *Code development (and use)*; *Patent searching*
advantages as means of organizing material, 532-533, 535
American Medical Association's Standard Nomenclature of Diseases, 244
American Society of Anesthesiologists, system in punched card file, 168
American Society for Metals-Special Libraries Association Metallurgical Literature Classification, 100-124, 248-260
for books, 529-530, 544
and characteristic properties, 530-532, 539-540
chemical classification code for Wyandotte-ASTM cards, 187-205
for chemical compounds and their coding, 187-205, 266-267, 465-468
and code development, 244, 530-532
consistent use of terminology, 538.
See also *Terminology selection*.
definition of terminology, 538-539
design determined by purpose, 262-263, 532, 535-536
determining main divisions, 537
and future expansion, 537-538
inadequacies for U. S. Patent Office, 267-268
for inorganic chemistry, 326-327, 597-602
introductory theoretical discussion, 528-533
of machines and systems, 8-9
for metallurgy, 102-105, 118-124
and modulation of subdivisions, 538
and mutual exclusiveness of terminology, 539
provision for expansion, 111-112
for quick discard of unwanted material, 535
relation to alphabetical indexing, 533-534
relation to punched-card code, 528-532
role of cross-references, 539
role of miscellaneous headings, 537-538
role of predicables, 539-540
schedules printed from punched cards, 565
and selection of terminology, 537-538
structure determined by purpose, 262-263, 532, 535-536
for technical reports indexed by Uniterm system, 156
in U. S. Patent Office, 261-267
- Coal reserves survey, with IBM card system, 320-321
- Code development (and use). See also *Punching schemes*; *Subject analysis*; *Terminology selection*; *Word coding*
advantages of simple codes, 27-28, 414-416
for aeronautical science, 344-350
for aluminum literature, 112, 114-116

- for anesthesia records, 163-173
- for broad fields, problems raised, 511-513
- for chemical compounds. See *Chemical compounds*
- and classification systems, 187-205, 244, 250-252, 528-532
- comparison of different punching schemes, 401-405, 462-463
- complex codes, as machine language, 396
- descriptors for Zatocoding system, 346-350
- establishing rules and procedures, 396-398
- and extra cards. See *Extra cards*
- and generic terminology, 242-243, 250-252, 269-270
- for Gmelin Information Center, 596-603
- for Hanawalt groups, in X-ray diffraction powder data, 208-213
- improvements in Keysort file, based on experience, 309-311
- and indication of relationships between concepts, 250-252, 399-402, 463-464, 494
- for inorganic chemistry, 597-602
- introductory and general discussion, 4-11, 18-19, 22-23, 27-28, 391-395
- with letter use frequencies, 497-504
- for metallurgical literature system, 102-111, 112, 114-116, 250-252
- for organosilicon chemistry file, 93-99
- principle of constructing codes, 406-418
- purpose of user as controlling factor in, 349, 414-416, 419-420, 532, 535-536
- with randomizing squares, 492-509
- and revision of Universal Decimal classification, 375-379
- for specialized files on punched cards, 303-339
- for spectral data, 178-183, 185-214, 217-231
- synonymous terms and related problems, 241-242, 250-252, 374-377, 404-405
- Coden system for coding journal references, 325
- Coffee literature bibliography, with punched card system, 335-336
- Collators, IBM systems, 58, 206, 367-368
- Collators, Remington-Rand system, 69-70
- Color identification of cards, to facilitate filing or searching, 245-246, 297-299, 332, 383
- Combination coding, values assigned to holes and punches, 425
- Combinations of concepts. See *Code development (and use)*; *Logical relations between concepts*; *Sorting operations*
- Compendia and encyclopedias. See also *Gmelin Handbuch*
- historical notes, 578-579
- in searching chemical literature, 548-549
- Computation (operations and problems). See also *Accounting*
- and data correlation, 432-437
- with IBM machines, 59-64, 319, 321
- introductory and general discussion, 10-11
- of letter use frequencies, 498-502, 505
- with Remington-Rand machines, 67-68, 69-72
- of word frequencies, in linguistic analysis, 361
- Computers. See *Data processing equipment*
- Concept analysis and coding. See *Code development (and use)*; *Subject analysis*
- Concordance, preparation of, with punched cards, 360-366
- Consonant code, word coding system, 497-498
- Converse (negative) searching, 146, 185-187, 195, 202, 217, 236-238, 278
- Copyflex copying, and edge-punched or hand-sorted cards, 560-561
- Copying methods. See also *Edge-punched cards*; *Photographic copying*; and specific systems, such as *Copyflex*; *Facsimile*; *Mimeograph*; *Ozalid*; *Verifax*; etc.
- for current distribution of information service, 135
- general review, 555-568
- "Microcite" system, 147-149

- role in establishing file, 135, 382-384, 556
- Correlation of data. See also *Sorting operations*
 - advantages of using punched cards, 175-176, 433
 - in biological tests file, 305
 - for chemical compounds, 466-467, 471, 475
 - introductory and general discussion, 7, 10-11, 432-433, 466-467
 - in organosilicon file, 98
 - and plant breeding and genetics, 386
 - in spectral data files, 175-177, 195, 202
- Corrosion data, with punched card system, 328-329
- Cost of books index (proposed), prepared with punched cards, 300-301
- Counters, electric, for edge-punched cards, 41
- Crime reports file, with punched card system, 336-337
- Cross-referencing
 - in classification systems, 539
 - in indexing, 517-518
 - in Uniterm system, 158-160
- Customers' service school
 - for IBM system, 54
 - for Remington-Rand system, 64
- Data processing equipment
 - with chemical compound coding, 466, 482, 489-491
 - and copying and transcription methods, 567-568
 - general description, 619-629
 - IBM systems, 62-64, 252-253
 - input devices, 620-621
 - output devices, 621-622
 - possible use in information systems, 89-90, 252-253, 405-406, 629-635
 - Remington-Rand system, 71-72
 - SEAC, used for patent searching, 275, 278
 - storage devices, 622-624
- Date, coding of
 - on edge-punched cards, 20, 96, 331
 - with Powers-Samas cards, 380-381
- Debugging, with data processing equipment, 634-635
- Decimal notation code. See also *Numerical coding*; *Universal Decimal classification*
 - for biological files, 308-309
 - for organic compounds, 468-476
 - for photographic data files, 311
 - for plant breeding and genetics, 375-379
- Deep-punching in edge-punched cards, 17-18. See also *Double holes in edge-punched cards*
- Descriptors for Zatocoding system. See *Code development (and use)*
- Digital computers. See *Data processing equipment*.
- Direct coding, values assigned to holes and punches, 19, 425
- Direct coding of edge-punched cards. See also *Punching schemes*
 - combined with other types of codes, 106-110, 446-447
 - and extra cards, 438
 - introductory and general discussion, 19
 - mathematical analysis, 439-440
 - for metallurgy, 106, 108, 110
 - in organosilicon compounds file, 94
 - and sequence sorting, 439-440
- Document card indexing systems. See *Edge-punched cards*; *Hand-sorted cards (not edge-punched)*; *IBM cards (and machines)*; *Remington-Rand cards (and machines)*
- Documentation, scope and definition, 575-577
- Double holes in edge-punched cards, punching and sorting operations, 17-18, 21
- Dow (Opler and Norton) code for organic compounds, 480-482
- Dropping fraction
 - actual and predicted, 460-463
 - definition, 446-447
 - and direct code combinations, 446-447
 - and overlapping effect, 453-455
 - in superimposed coding, 447-460
- Dry mounting. See also *Edgepunched cards*; *Mounting of materials*
 - of abstracts, data, etc., on edge-punched cards, 27

- Duplicating machines and methods. See *Copying methods*; *Grooving machines*; *Photographic copying*; *Reproducing punches*.
- Dyson code for organic compounds, 476-479
- Economic information, searches for, 552-553
- Edge-punched cards. See also *Punched cards, design and layout*
and Addressograph copying, 558
adhesives, suitable for attaching clip-pings, 27
as aids in literature searching, 99
alignment block for sorting, 32-33, 46-47
for anesthesia records, 161-174
and bibliography preparation, 6-7, 10-11, 19, 25, 98-99, 561-562
for biological data files, 305, 307-311
for chemical data files, 324-329
classification of, 8
for coffee literature file, 335-336
copying and transcription methods, 558-562
and Copyflex (diazotype) copying, 560-561
correction of punching errors, 31-32, 38
for correlation of data, 98
direct sorting, 13-17
dry-mounting of abstracts, 27
duplicating notches, methods for, 558
electric counter, 41
electric keypunch, 33-34, 40, 44-45, 49-50
for electron microscopy file, 335
and facsimile copying methods, 559-560, 561-562
film transparencies used as, 319-320
foreign manufacturers, 53
grooving machines, 34-35, 40-42, 44-45
hand punches, 32-35, 39, 45, 52
and hectograph copying, 558
for hobbies files, 333-334
improvements of system, based on experience, 309-311
in inorganic chemistry, 607
introductory and general description, 4-5, 12-29
for laboratory records files, 312-314
in legal data file, 322-323
for library routines, 281-283, 285, 290-292, 296-299
in manuscript preparation, 93-99
for market research file, 315-318
for medical reports (deep-sea diving accidents), 336-338
for metallurgical literature, 105-111
with microcards mounted, 559
with microfilm insert, 31, 74-75, 77, 559
with microtape mounted, 82, 559
and mimeograph or multilith copying, 558-559
and mounting of graphical materials, 27, 99, 560
for nondestructive testing literature file, 338-339
for nuclear data files, 303-305
for nutritional research file, 337-338
for organosilicon compounds, 96
and Ozalid copying, 31, 39, 560-561
for personnel classification, 335
for petroleum data files, 329-331, 333
and photographic copying methods, 558-562
for photographic data files, 311-312
and Polaroid camera copying, 561
precautions during filing and use, 27-28
prepared from ordinary file cards, 49-51
provision for expansion of codes, 111
and reflex photocopying, 560
review of commercial types, 30-53
sorting devices, 32-36, 39-42, 46, 52
sorting rate, 32, 36, 40-41, 52
with special printing. See *Punched cards, design and layout*
for spectral data files, 176, 178-183, 209-211, 215-216, 223-225
storage cabinets, 36, 46
and superimposed coding for organo-silicon compounds, 94-95
in technical services files, 314-315
and Thermofax copying, 560
and Verifax copying, 560
and Xerographic copying, 559, 561
"Ekaha" system, based on peek-a-boo principle, 127

- ELCO code, word coding system, 498-499
- Electrofax enlarger printer, for use with microfilm inserts, 559
- Electron microscopy bibliography, with punched card system, 335
- Electronic calculator, IBM system, 62
- Electronic statistical machine (Type 101), IBM system, 63, 207, 236-269, 273, 367-368, 508
- Enlargers for microfilm inserts, 78-80
- Equipment capabilities, and sorting operations, 398-406
- Errors in punching
 - correction in edge-punched cards, 31-32, 38
 - correction in hand-sorted cards (not edge-punched), 49
 - correction in peek-a-boo cards, 139
 - detection in IBM cards, 56-57
 - detection in Remington-Rand cards, 66
- Explosives data, with punched card system, 335
- Extra cards
 - and direct coding, 438
 - from generic terms in code, 438
 - mathematical analysis and superimposed word coding, 447-462
 - from mechanical sorting operations, 233-234, 438
 - in numerical coding, 442, 444
 - in organosilicon chemistry file, 95
 - from small vocabulary in peek-a-boo system, 144-145
 - and superimposed coding, 233-234, 345-346, 401-402, 425, 438-439, 494-495
 - in unit card systems, 144-145
 - and Zatocoding, 345-346
- E-Z sort system. See also *Edge-punched cards*; *Punched cards, design and layout*
 - in biological data file, 309-310
 - cards and equipment, 37-42
 - card savers, 38
 - for coffee literature file, 335-336
 - electric counter, 41
 - in legal data file, 322-323
 - in market research file, 315-318
 - for metallurgical literature system, 105-111
 - multi-sorter, 40-41
 - in photographic negatives file, 312
 - punches, manually and electrically operated, 39-42
 - sorting equipment, 39
 - specially designed cards, for word coding methods, 39
- Facsimile reproducers, and edge-punched or hand-sorted cards, 559-560, 561-562
- File arrangement, preparatory to sorting, 12-13, 18
- File cards, ordinary, conversion to punched cards, 28-29, 49-51
- Filing cards in groups
 - aided by color identification, 245-246, 297-299, 313, 332, 383-384
 - to facilitate searching, 134-135, 214-215, 245-246, 332, 384
- Filing punched cards
 - cabinets for edge-punched cards, 36, 46
 - cabinets for hand-sorted cards (not edge-punched), 47, 49
 - precautions recommended, 27-28
- Filing system, numerical (Radex), 87
- Film storage and selecting system (Filmorex system), 88-89
- Film transparencies, used as edge-punched cards, 319-320
- Filmorex system, 88-89
- Filmsort system. See also *Edge-punched cards*; *Punched cards, design and layout*
 - cards and equipment, 74-81
 - duplicating equipment, 79, 81
 - for engineering blueprints file, 338
 - enlarging and reading equipment, 78-80
- Filmsnips, 77
- Mounter, 77
 - photo transparencies mounted in E-Z sort cards, 312
- Findex system
 - cards and equipment, 47-49
 - filing cabinet, 47, 49
 - manual punch and sorting equipment, 47-49
 - selector, 47, 49

- Fine sorting of edge-punched cards, 24-25
- Flexisort system
cards and equipment, 49-51
electrically operated punch, 49-51
- Frear code for chemical compounds, 468-470
- Frequencies of letters, basis for coding, 498-502, 505
- Frequencies of words in text, analysis of, 361, 373
- Fuels and lubricants additives, with punched card system, 331-333
- Geochemistry data, with punched card system, 326-328
- Geological data, with punched card systems, 319-321
- GKD code for chemical compounds, 487-489
- Gmelin Handbuch
arrangement by system number, 588-590, 593-596
basis for information center, 592, 595-596
historical notes, 579, 581, 587, 590
organization of, 588-590, 593-596
present-day problems and status, 606
scope of field covered, 587-588
- Gmelin Institut
future trends, 612
and mechanized documentation, 585-612
organization plan, 609-611
and punched-card techniques, 596-609
and subject archives, 592-593
- Golden hamster, indexed on edge-punched cards, 307-308
- Gordon-Kendall-Davison code for chemical compounds, 487-489
- Government publications (in literature searching), 552
- Graphic output, in data processing equipment, 621-622
- Graphical materials mounted on edge-punched cards, 99
- Grooving machines for edge-punched cards, 34-35, 40-42, 44-45
- Gruber code for organic compounds, 479
- Guides to patents in literature searching, 549-551
- Hanawalt Groups code, for X-ray diffraction powder data files, 208-213
- Handbooks and compendia in searching chemical literature, 548-549
- Hand-sorted cards. See *Edge-punched cards*; *Hand-sorted cards (not edge-punched)*; *Unit card systems (hand manipulated)*
- Hand-sorted cards (not edge-punched)
classification of, 8
copying and transcription methods 558-562
correction of punching errors, 49
Findex system, 47-49
with microtape mounted, 82
sorting devices, 47-49
sorting rate, 47
storage cabinets, 47, 49
- Hectograph copying, and edge-punched or hand-sorted cards, 558
- Hobbies, with punched card systems, 333-334
- Holes in cards
definition of different types, 422-423
values assigned to, 424-426
- Hydrocarbons data and properties, coded on IBM cards, 331
- Hollerith cards. See *IBM cards (and machines)*
- IBM** cards (and machines)
in accounting, 54-64
accounting machine, 59-60, 368-369
alphabetical coding, 6, 54-55, 494-504
for astronomical data files, 321
for biological data files, 305-306, 311
calculating punches, 61-62
card combined with microcards, 323-324
card as hand-sorted cards, 226, 303
card-operated typewriter, 565-566
card-programmed electronic calculator, 62
Cardatype, 361, 563, 565, 566
for cataloging in libraries, 285-287
for chemical data file, 325-326
for circulation control in libraries, 288, 290-296
for circulation control of talking book machines, 294-295
coding capacity of cards, 54, 184-185

- and coding of inorganic compounds, 597-605
 - collators, 58, 206, 367-368
 - computers system, for intern matching program, 335
 - computing applications, general discussion, 59-64
 - for crime reports file, 336-337
 - customers' service school, 54
 - data processing machines, 62-64
 - dial board for 101 machine, 238-239
 - electronic calculator, 62
 - electronic statistical machine (Type 101), 63, 207, 236-239, 273, 367-368, 508
 - experimental machine X-794, 252-253
 - for explosives data file, 335
 - for geological data file, 320-321
 - for hydrocarbon data files, 331
 - introductory and general discussion, 3-6, 54-64
 - keypunch, 55-56
 - in linguistic analysis, 358-369
 - for manuscript preparation, 562, 564-566
 - mark-sensing, 56, 236, 244, 364
 - for meteorological data files, 318-319
 - microcards fitted into IBM cards, 323, 324
 - for nuclear data file, 303
 - numerical coding, 6, 54-55, 505-508
 - for ordering routines in libraries, 280-281, 283-284
 - for printing catalogs, indexes, lists, tables, etc., 285-286, 300-301, 562, 564-566
 - printing punch, 563
 - printing units, 59-60, 62-64, 564-566
 - punching machines, 55-57, 60-62, 563
 - reproducing punches, 60-61, 563
 - service bureaus, 54
 - sorters and sorting, 57-58, 63, 184-186, 202, 206-207, 215, 217-218, 236-237, 367
 - sorting rates, 57-58, 63, 237, 321, 361
 - for spectral data files, 176, 183-195, 202-207, 211-215, 216-223, 225-230
 - for statistical analyses in libraries, 299-300
 - summary punch, 60, 367
 - tabulators, 564-565
 - transcribing punching to tape, 563, 566
 - typewriter card punch, 56
 - typing information on cards, 562, 566-567
 - verifying machine, 56-57
- Indexes
- assembly of, 520-521
 - and book classification, 543
 - for broad fields, problems raised, 511-513
 - and classification systems, 533-534
 - and coding of chemical compounds, 465-468
 - correlative, in book form, 510-511
 - correlative, of broad fields, and problems raised, 511
 - and cross references, 517-518
 - different types, 513, 631-632
 - editing and arranging, 521, 546
 - facilitating use of, 513-520, 544
 - indirect approach when searching, 546-547
 - introductory and general discussion, 510-513
 - for large-scale retrieval systems, 631-632
 - manipulative and non-manipulative, 510-513
 - and modifications for subject headings, 520
 - preparation (operations and principles), 130-134, 246, 513-522, 524
 - preparation with punched cards (operations and principles), 387, 562, 564-566
 - qualifications for preparing, 515, 521-522
 - and scientific abstract periodicals, 387, 510-524, 546
 - special types, prepared with data processing equipment, 372-373
 - special types, prepared with punched cards, 360-366
 - structure of, 518-519
 - synonyms and related problems, 131-132, 515-518
 - technique of searching, 522-524
 - and terminology problems, 131-132
 - words as tools in constructing, 515-518

Information service

- in chemistry, future possibilities, 512-513
 - copying and transcription problems, 555-556
 - for corrosion data, 328-329
 - for electron microscopy bibliography, 335
 - expedited by group-filing of cards, 384
 - in geology, on film transparencies, 319-320
 - for hydrocarbons data and properties, 331
 - in inorganic chemistry, 606-607
 - in metallurgy, 100-102, 248-250
 - for nondestructive testing literature, 338-339
 - in organosilicon field, 98-99
 - in plant breeding and genetics, 384-387
 - using punched cards for standing inquiries, 386-387
- Infrared absorption spectra**
coded on punched cards, 177-207
indexed by peek-a-boo methods, 146-147
- Input devices for data processing machines, 620-621**
- Insecticide effectiveness, correlation with punched cards, 434-435**
- Interfixes, in Patent Office coding system. See *Chemical Compounds***
- Intermediate punching in edge-punched cards, 17-18**
- International Business Machines Corp. See *IBM cards (and machines)***
- International Union of Pure and Applied Chemistry (IUPAC), study of chemical coding, 467-468**
- Intern matching program, 335**
- Interpreting punched cards**
IBM system, 563-564
Remington-Rand system, 68, 563-564
- Ion mass, coded on punched cards, 225**
- Justowriter equipment for automatic copying process, 562**

Keypunch

- for IBM cards, 55-56
 - for Remington-Rand cards, 64-66
- Keysort system. See also *Edge-punched***

cards; Punched cards, design and layout.

- for anesthesia records, 161-174
 - for bindery records, in libraries, 285
 - in biological data files, 305-310
 - cards and equipment, 30-37
 - card savers, 31-32
 - in chemical data files, 324-329
 - for circulation control in libraries, 290-292, 297-299
 - data punch, 37
 - for deep-sea diving accidents files, 336-338
 - for electron microscopy file, 335
 - filing cabinet, 36
 - for hobbies files, 333-334
 - in laboratory records files, 312-314
 - microfilm inserts, 31, 75, 77
 - multiple needle selector, 34-36
 - in nuclear data files, 303-305
 - for ordering routines in libraries, 281-282
 - and Ozalid copying methods, 31
 - for personnel classification, 335
 - in petroleum data files, 329-331, 333
 - in photographic data files, 311-312
 - punches, manually and electrically operated, 32-37
 - selector, for hand sorting, 34-36
 - service bureaus, 30
 - sorting equipment, 32-36
 - for spectral data files, 178-183, 209-211, 215-216, 223-225
 - tabulating punch, 36
- Laboratory records, with punched card systems, 312-314**
- Legal data, with punched card systems, 322-323**
- Library routines**
application of punched card systems, 279-302
bindery records, with edge-punched card systems, 285
book budget accounting, with machine-sorted card system, 281
cataloging, with machine-sorted card systems, 285-287
circulation control
with edge-punched card systems, 290-291, 296-299

- with machine-sorted card systems, 288-297
- ordering
 - with edge-punched card systems, 281-283
 - with machine-sorted card systems, 280-281, 283-284
- shelf-listing, with machine-sorted card system, 281
- Linguistic analysis
 - with aid of data processing equipment, 372-373
 - with aid of punched cards, 357-372
- Listomatic camera for automatic copying process, 562
- Literature searching. See also *Abstract periodicals*; *Classification systems*; *Indexes*; *Patent searching*; *Sorting operations*
 - and American library resources, 547-548
 - and chance discoveries, 552
 - indirect approach to information, 546-547
 - in inorganic chemistry, 571-612
 - for market research, 552-553
 - in metallurgy, 116, 256-260
 - methodical techniques, 99, 246-247, 256-260, 355-356, 547-548, 552-554
 - Patent office requirements, 265-267
 - and patent validity, 262-263, 265-267
 - preliminary orientation, 545
 - products of the search, 132-134, 238-239, 246-247, 260
 - and research planning, 542-543, 582
 - sources and methods, 542-554, 577-585
 - use of indexes, 522-524
 - and utilization of terminology, 256-260, 545-546
- Logic and punched card systems, theory and analysis, 431
- Logical analysis of punched cards, 426-429
- Logical combinations of punching positions, 426-429
- Logical operations applicable to punched card systems, 426-429
- Logical relations between concepts. See also *Code development (and use)*; *Sorting operations*
 - specified for searching, 237-239, 254-256, 269-270, 271, 276-277, 399-402, 426-429, 494
- Machine language (complex coding schemes), 396
- Machine-sorted cards. See *IBM cards (and machines)*; *Powers-Samas cards*; *Punched cards*; *Remington-Rand cards (and machines)*; *Samas (Underwood) cards*; *Unit card systems (machine-manipulated)*
- Machine-sorted card systems, classification of, 8-9
- Magnetic cores, discs, drums and tapes, in data processing equipment, 621-624
- Major-minor coding, 458, 460-461. See also *Punching schemes*
 - of organosilicon compounds, 94-95
- Manuscript preparation
 - by card-operated typing devices, 565-566
 - with edge-punched cards, 93-99
 - use of information entered on cards, 96-99
- Mark sensing of IBM cards, 56, 236, 244, 364
- Market research data, with punched card systems, 315-318
- Mass spectra, coded on punched cards, 223-226
- Mathematical analysis
 - of alphabetical coding, 440-443
 - of coding systems, 408-414, 438-464
 - of direct coding, 439-440
 - of major-minor coding, 440-443
 - of numerical coding, 440-443
 - predictions and actual results, 460-462
 - of selecting operations directed to combinations of concepts, 408-414
 - of selector codes, 440-443
 - of sequence codes, 443-445
 - of superimposed coding, 447-464
 - of triangular coding, 441-442
- Mathematical tables, preparation with punched cards, 319, 321, 565
- McBee cards. See *Keysort system*
- Mechanization of literature searching, needs and possibilities, 391-393, 571-573, 582-585

- Medical data. See also *Anesthesia records*
correlation using punched-card techniques, 233, 436
deep-sea diving accidents, with punched card system, 336-338
- Medical students (interns) matching program, with computer system, 335
- Medicinal compositions, in Patent Office mechanized searching test, 268-269
- Melting point, coded on punched cards, 178, 205, 214
- Metallurgy. See also *Aluminum and aluminum alloys*; *Metals and alloys*
abstracting system, on edge-punched cards, 112, 114-117
classification system for, 102-105, 118-124
code for common variables index, 105, 110-111
code for materials index, 104-105, 108-110
code for processes and properties index, 103-104, 106-108
information file, on edge-punched cards, 112, 114-117
- Metals and alloys. See also *Aluminum and aluminum alloys*; *Metallurgy*
coded on edge-punched cards, 108-110
grouped for coding, 109-110
- Metals plating data, with punched card system, 324-325
- Meteorological data, with punched card systems, 318-319
- Mice, inbred strains of, indexed on Keysort cards, 305-306
- Microcards
combined with IBM cards, 323-324
mounted on edge-punched cards, 559
"Microcite" system of copying equipment and operations, 147-149
linked to peek-a-boo system, 147-149
- Microfilm, mounted on edge-punched cards, 318
- Microfilm inserts
cards for, 74-77
duplicating equipment for, 79, 81
in edge-punched cards, 31, 74-75, 559
for engineering blueprints file, 338
enlarging and reading equipment for, 78-80, 559
Filmsort system, 74-81, 559
in machine-sorted cards, 74-75, 338, 559
Ozalid copying of, 81
- Microtape, for mounting microfilm on cards, 81-82, 559
- Microtape readers, 82
- Micro-typewriter for typing on IBM cards, 566-567
- Mimeograph copying, and edge-punched or hand-sorted cards, 558
- Minicard equipment, 252-253
- Modification of index subject headings, 520
- Modulants, in Patent Office coding system. See *Logical relations between concepts*
- Modulation of subdivisions in classification, 538
- Molecular weight, coded on punched cards, 224
- Mounting of material on cards. See also *Adhesives*; *Dry mounting*; *Edge-punched cards*. 27, 82, 99, 352, 559
- Multilith copying, and edge-punched or hand-sorted cards, 558, 559
- Multiple needle sorting
devices for, 34-36, 40-42
and superimposed word coding, 455
- Multiple sequence sorting, of edge-punched cards, 25-26
- Name coding. See *Author names*
- National Bureau of Standards, instrumentation indexing and searching system, 129-130, 136-147
- National Bureau of Standards, punched card system for spectral data, 178-183
- National Research Council, code for chemical compounds, 470-476
- National Research Council Committee on Spectral Absorption Data, punched card system, 178-183, 215-216
- Needlesort system, cards and equipment, 51-52
- Negative searching. See *Converse searching*
- Nodal index code for chemical compounds, 491
- Nomenclature problems. See *Code development (and use)*; *Terminology selection*

- Nondestructive testing (NDT) literature, with punched card system, 338-339
- Nuclear data files, with punched card systems, 303-305
- Numerical coding
 based on binary numbers, 441-445
 combined with other types of codes, 445-446, 505-507
 for data processing equipment, 624-625
 for edge-punched cards, 19-20
 and extra cards, 442, 444
 with IBM cards, 6, 54-55, 505-508
 mathematical analysis, 440-443
 in peek-a-boo system, 146
 with Remington-Rand cards, 6, 64-65
 with self-checking features, 505-507
 and sequence sorting of edge-punched cards, 24-25
 and superimposed word coding, 505-507
- Nutritional research, with punched card system, 337-338
- "Open-endedness" of file, general discussion, 135-136
- Opler and Norton (Dow) code for organic compounds, 480-482
- Optical coincidence subject cards. See *Peek-a-boo system*; *Unit Card systems*
- Ordering routines in libraries
 use of edge-punched cards, 281-283
 use of IBM cards, 280-281, 283-284
- Organic compounds. See *Chemical compounds*
- Organosilicon compounds. See *Chemical compounds*
- Orthodontic diagnostic data, on Keysort cards, 307
- Output devices for data processing machines, 621-622
- Ozalid copying
 and edge-punched or hand-sorted cards, 31, 39, 560-561
 with microfilm inserts, 81
- Parasitology file, indexed on edge-punched cards, 308
- Patent Office. See *U. S. Patent Office*
- Patent searching. See also *Classification systems*; *Literature searching*; *Sorting operations*
 for alternative disclosures, 277
 classification system, conventional type, 261-268
 different types, 262-263, 265-267, 551
 and examination of applications in U. S. Patent Office, 262-263, 265-267, 533
 experimental mechanized systems, 272-273, 278
 and patent interpretation and validity, 262-263, 265-267, 549-550
 problems, based on search requirements, 267-270
- Peek-a-boo system
 applicable to large collections of documents, 149-151
 cards and equipment, 137-141, 147-149
 correction of punching errors, 139
 electronic equipment systems, 150-151
 foreign manufacturers, 127
 for indexing infrared absorption spectra, 146-147
 for instrumentation literature, 129-130, 136-147
 for inorganic chemistry, 607-609
 introductory and general discussion, 125-141
 with "Microcite" system of copying, to prepare bibliographies, 147-149
 punching devices and operations, 137-140
 sorting and read-out devices and operations, 140-141
 terminology selection and categorization, 142-145
- Personnel classification, with punched card system, 335
- Pesticides information, with IBM card system, 311
- Petroleum data, with punched card systems, 329-333
- Photographic copying
 for bibliographies prepared with punched cards, 561-562
 diazotype photography, 560-561
 and edge-punched or hand-sorted cards, 558-562

- Photographic data files, with punched card systems, 311-312
- Photographic slides
indexed with punched card system, 334
used as edge-punched card system, 334
- Photographic storage devices
classification of, 9
for data processing equipment, 624
- Physical properties of chemical compounds. See also specific properties, as *Boiling Point*; *Melting point*; *Ultraviolet absorption spectra*; etc.
coded on punched cards, 178, 185-187, 205, 212-214, 217-219, 221-225
- Plant breeding and genetics file
card design and layout, 381-382, 385
code development and terminology problems, 374-377
filing cards in groups to facilitate search, 384
and information service, 384-387
and Powers-Samas system, 379-386
and preparation of indexes, 387
procedure in coding papers, 379-380
punching operations, 381-384
and search to combinations of index entries, 385-386
and Universal Decimal Classification, 375, 378-386
- Polaroid Land camera copying, and edge-punched or hand-sorted cards, 561
- Posting machine, Remington-Rand system, 69
- Powers-Samas cards, for plant breeding and genetics, 374-387
- Predicables, role in classification, 539-540
- Printing devices
in data processing equipment, 621
general review, 555-568
IBM system, 59-60, 62-64, 564-566
operated by magnetic tape, 567-568, 621
Remington-Rand system, 67-68, 564
- Programming
of data processing equipment, 627-629
different types, 627-629
of WRU Searching Selector, 256-260
- Publication, growth and problems, 577-582
- Punched cards. See also *Edge-punched cards*; *IBM cards (and machines)*; *Remington-Rand cards (and machines)*; *Punched cards, design and layout*
in data processing equipment, 620-621
holes and punches defined, 422-423
information content defined in binary code, 423-424
values assigned to holes and punches, 424-426
- Punched cards, design and layout. See also *Code development (and use)*
edge-punched cards
for anesthesia records, 163, 172
for biological data, 306-307, 309-310
for chemical data files, 324, 329
for coffee literature survey, 336
for deep-sea diving accidents, 338
for laboratory records, 314
for legal data file, 323
in library routines, 282, 297, 298
for market research file, 317-318
for metallurgical literature, 105-111
for organosilicon chemistry, 96
provisions for expansion of code, 111
for spectral data, 179, 209, 224
for technical services file, 316
- IBM cards
for biological-medical data file, 245
for chemical data file, 325
codes punched in horizontal rows, 271-272
coding capacity, 54
for crime reports, 337
for inorganic chemistry (Gmelin system), 604-605
in library routines, 281, 283, 286, 291, 292-293, 295, 296, 301
in linguistic analysis, 360, 362
mark-sensing, for biological-medical data file, 236
for patent searching, 272
with randomizing square, 508
for spectral data, 185, 212, 217, 221, 227
for superimposed word codes, 508
- peek-a-boo cards
for inorganic chemistry, 609
for instrumentation literature, 126, 128-129, 137
- Powers-Samas cards, for plant breeding and genetics, 381-382, 385

- Remington-Rand cards, coding capacity, 64
- Samas (Underwood) cards, in library routines, 290
- Zatocard, for aeronautical sciences file, 345
- Punches in cards
- definition of different types, 4-6, 422-423
 - values assigned to, 19-23, 424-426
- Punching devices, machines and operations
- for converting ordinary file cards to punched cards, 28-29, 49-51
 - for edge-punched cards, 30-53, 352-354
 - general and introductory discussion, 3-6, 27
 - for hand-sorted cards (not edge-punched), 47-49
 - for IBM cards, 55-57, 60-62
 - for peek-a-boo system, 137-140
 - for Powers-Samas cards, 381-384
 - and preliminary marking of cards, 27, 352-354
 - rate of punching, in peek-a-boo system, 140
 - for Remington-Rand cards, 64-66, 69-74
 - for Zatocoding system, 352-354
- Punching schemes
- comparative evaluation and development, 462-463
 - mathematical analysis, 438-464
- Pyramid codes, for edge-punched cards, 21, 441-442
- Radex filing system, 87
- Radioactive materials, identification with Keysort cards, 304
- Radioisotopes, identification with Keysort cards, 304-305
- RAM, storage device for data processing equipment, 624
- Random codes, prepared with randomizing squares, 492-509
- Random number coding. See also *Scatter coding*; *Superimposed coding*
- for punched card files, 232-247, 344-346, 401-402, 505-507
- Rapid selector, and copying and transcription methods, 567
- Readers
- for microfilm inserts, 78-80
 - for microtape, 82
- Reflex photocopying, and edge-punched or hand-sorted cards, 560
- Remington-Rand cards (and machines).
- See also *Powers-Samas cards*
 - in accounting, 64-74
 - alphabetical coding, 6, 64-65
 - automatic punch, 64-66
 - calculating punch, 70-71
 - cards used as hand-sorted cards, 128
 - for cataloging in libraries, 287
 - coding capacity of cards, 64
 - collating reproducer, 69
 - collators, 69-70
 - computer, 71-72
 - computing applications, general discussion, 67-68, 69-72
 - customers' service schools, 64
 - electronic sorter, 66-67
 - interfiling reproducing punch, 69
 - introductory and general discussion, 3-6, 64-74
 - for legal data files, 322
 - multi-control reproducing punch, 69, 563
 - numerical coding, 6, 64-65
 - for petroleum data file, 331-333
 - posting interpreter, 68
 - posting machine, 69
 - printing machines, 67-68, 564
 - punched card interpreters, 68
 - punching machines, 64-66, 69-74, 563
 - reproducing machines, 69-70, 73
 - service bureaus, 64
 - sorters and sorting, 66-67, 236-237
 - sorting rates, 66-67, 237
 - summary card punch, 72-73
 - Synchro-matic, 72
 - tabulators, 564
 - tag control reproducer, 73
 - tape recording and transcription, 73-74
 - verifying machine, 66
- Report literature, growth and problems, 581-582
- Reproducing punches
- IBM system, 60-61
 - Remington-Rand system, 69-70, 73

- Research planning
 with aid of edge-punched cards, 98
 and scientific information, 542-543, 582
- Samas (Underwood) cards
 cards as hand-sorted cards, 129
 for circulation control in libraries, 289-290
- Scatter coding (random-like patterns)
 for author names, 354-355
- SEAC (National Bureau of Standards computer). See *Data processing equipment*
- Searching operations. See *Literature searching*; *Sorting operations*
- Selectivity in searching, 234-235, 407-418
- "Selecto" system, based on peek-a-boo principle, 127
- Selector coding
 combined with other types of codes, 445-446
 for edge-punched cards, 20-21, 26-27
 mathematical analysis of, 440-443
- Self-demarking word code, 499
- Semantic code dictionary, principles and use, 252, 258-260
- Semantics. See *Code development (and use)*; *Terminology selection*; *Word coding*
- Sequence coding
 combined with other types of codes, 445-446
 mathematical analysis of, 443-445
 and sorting edge-punched cards into order, 23-26, 443-445
- Serial number,
 coded on edge-punched cards, 96
 coded on IBM cards, 205-206, 214, 220, 223, 231
- Service bureaus
 for IBM system, 54
 for Keysort system, 30
 for Remington-Rand system, 64
- Shallow punching in edge-punched cards, 17-18
- Shelf-listing in libraries, with IBM system, 281
- "Sichlochkarten" system, based on peek-a-boo principle, 127
- Significant letter code, word coding system, 498
- Soil analysis data, correlation with punched cards, 434
- Solvents for chemical compounds, coded on punched cards, 218-219
- Sorting devices
 for edge-punched cards, 32-36, 39-42, 46, 52
 for hand-sorted cards (not edge-punched) 47, 49
 IBM system, 57-58, 63, 184, 237, 367
 Remington-Rand system, 66-67, 237
- Sorting operations
 converse (negative) searching, 146, 185-187, 192, 202, 217, 236
 with data processing equipment, 405-406, 630-631
 directed to combinations of concepts, 155-156, 237-239, 254-256, 271, 276-277, 385-386, 391-392, 399-402, 426-429, 493-494. See also *Logical relations between concepts*
 with double holes in edge-punched cards, 17-18, 21
 and equipment capabilities, 398-406
 facilitated by filing in groups, 245-246, 384
 for hand-sorted cards (not edge-punched), 47, 49
 for IBM cards, 57-58, 63, 236-239, 245-247, 493-494, 508
 introductory and general description for edge-punched cards, 4-5, 12-27, 32-36, 39-42, 46, 52
 for unit card systems (hand manipulated), 132-135
 and manuscript preparation, 96-99
 preparatory positioning of edge-punched cards, 12-13, 18
 rate of
 with data processing equipment, 630
 with edge-punched cards, 32, 36, 40-41, 52
 with Filmorex system, 88
 with Findex cards, 47
 with hand-sorted cards for library circulation control, 297
 with IBM cards, 57-58, 63, 237, 321, 361

- with Remington-Rand cards, 66-67, 237
- with Zatocoding system, 343-344
- for Remington-Rand cards, 66-67, 237
- with search patterns of codes, 234, 345, 405
- with selector codes, 26-27
- with sequence codes, 23-27, 443-445
- Spectral data, coded for punched cards, 175-231
- Spectroscopic data, correlation with punched cards, 436
- "Sphinxo" system, based on peek-a-boo principle, 127, 607-609
- Stamp collections, indexed with punched card system, 333
- Statistical analysis
 - of anesthesia records, 161, 173
 - of astronomical data, with IBM cards, 321
 - of chemical literature, 543
 - of crime reports, 336
 - of letters, their number and distribution in a text, 367-368
 - of letter use frequencies, 497-502, 505
 - in libraries, using punched card systems, 299-300
 - of word frequencies, 361, 373
 - of word structures and their frequencies, 367
- Stencil preparation by card-operated typewriter, 566
- Storage devices for data processing machines, 622-624
- Subject analysis. See also *Code development (and use)*
 - basic operations, 393-395
 - and code construction, 391-421
 - introductory and general discussion, 18-19, 22-23, 27, 391-395
 - in operation of peek-a-boo system, 145
 - in operation of punched card system, 246
 - in operation of Uniterm system, 153, 157
 - in operation of Zatocoding system, 350-352
 - and preparation of abstracts, 249-252
- Subject archives of Gmelin Institut, 592-593
- Subject-card indexing systems. See *Unit card systems*
- Summary card punch, Remington-Rand system, 72-73
- Summary punch, IBM system, 60, 367
- Superimposed coding
 - for aeronautical sciences, 344-346
 - combined with major-minor classification, 458, 460
 - for edge-punched cards, 21-22, 344-346
 - efficient punching schemes, 233-234, 451-457, 496-506
 - and extra cards, 233-234, 345-346, 401-402, 425, 438-439, 494-495
 - introductory and general discussion, 21-22, 492-494
 - mathematical analysis of, 447-464
 - predictions and actual performance, 460-462
 - for medicine, pharmacology and allied fields, 239-244
 - for organosilicon compounds, 94-95
 - and overlapping effect, 234, 453-455
 - values assigned to holes and punches, 425
 - variations of, 457-460, 492-509
 - and word coding, 451-464, 493-504
- Zatocoding variety, 344-346, 457 (footnote)
- Symbolism, selection of, for coding, 368-369, 375
- Synchro-matic, Remington-Rand system, 72
- Synonymous terms. See also *Terminology selection*
 - and code development, 241-242, 374-377, 404-405
 - and cross-references in indexing, 242-243, 517-518
 - and cross-references in Uniterm system, 159-160
 - and indexing problems, 131-132, 242-243, 515-518
 - in literature searching, 241-242, 545-546
- System number arrangement of Gmelin Handbuch, 588-590, 593-596
- Talking book machines, circulation control with IBM cards, 294-295

- Tape, magnetic or paper, in data processing equipment, 621, 623
- Tape-reading devices, classification of, 9
- Tape recording and transcription
of IBM punched cards, 563, 565-566
of Remington-Rand cards, 73-74
of superimposed word codes, 508-509
- Teaching application of punched cards, 337-338
- Technical services data, with punched card systems, 314-315
- Telegraphic-style abstracts. See *Abstracts, encoded*
- Termatex system
based on peek-a-boo principle, 129
cards and equipment, 84-85
- Terminology selection. See also *Synonymous terms*
categorization of terms, 142-144, 351-352
and code development, 22-23, 241-243, 347-350, 377, 416, 463, 530-532
and conventional classification, 537-538
and conventional indexes, 515-518
and indexing problems, 131-132, 242-243
involving generic relationships, 131-132, 242-243, 250-252
for unit card system (hand-manipulated), 142-145
for Uniterm system, 155-156, 157-160
for Zatocoding system, 340-350
- Thermofax copying, and edge-punched or hand-sorted cards, 557, 560
- Transaction cards. See *Circulation control in libraries*
- Triangular codes, for edge-punched cards, 21, 441-442
- Typewriters
card-operated, 565-566
controlled by magnetic tape, 568
micro, for typing on IBM cards, 566-567
- Typewriter card punch, IBM system, 56
- Ultraviolet absorption spectra, coded on punched cards, 215-220
- Union catalog of serials (proposed) prepared with punched cards, 300-301
- Unisort system
alignment discs, 45
card holder for punching machine, 46
cards and equipment, 43-46
filing cabinets, 46
manual punch and sorting equipment, 44-46
manually and electrically operated punches, 44-45
pull tubs, 46
sorting pan (alignment block), 46-47
specially printed cards, 43
- Unit card system (hand-manipulated)
number matching systems, 82-84
pattern coincidence systems. See *Peek-a-boo system*
- Uniterm system, 82-84, 152-160
- Unit card systems (machine-manipulated)
Alpha-matrex system, 85-88
electronic systems for large collections of documents, 150-151
inadequacies for U. S. Patent Office, 268
Matrex (matrix-index) systems, 84-88
pattern coincidence systems, 84-88
Termatex system, 84-85
- Uniterm system
card for, 82-83
card design and layout, 153-154
depth of indexing, compared with conventional system, 156
for indexing technical reports, 152-160
limiting size of file, 157
operating procedures, 153-156
in petroleum processing file, 152-160
rate of indexing, compared with conventional system, 156
selection of terminology, 155-156, 157-160
- Uniterm-Book, 83-84
- Universal Decimal Classification, revised for punched-card code, 375-379
- U. S. Patent Office code for chemical compounds, 270-274, 489-490
- Varityper equipment, for automatic copying process, 562
- Venn diagrams, used in logical analysis of punched cards, 412, 426-427

- Verifax copying, and edge-punched or hand-sorted cards, 560
- Verification of punching. See also *Errors in punching*
- IBM system, 56-57
 - Remington-Rand system, 66
- "Vicref" (Visual Cross Reference) system, based on peek-a-boo principle, 128
- Viewers for microfilm inserts, 78-80
- Visible absorption spectra, coded on punched cards, 220-223
- Vision in invertebrate animals, file on Keysort cards, 308-311
- W**estern Reserve University experimental equipment. See *WRU Searching Selector*
- Wiswesser code for organic compounds, 482-485
- Wood, data on, correlated with punched cards, 433-434
- Word coding
- consonant code system, 497-498
 - ELCO code system, 498-499
 - self-demarcating code system, 499
 - significant letter code systems, 498
 - and superimposed coding, 451-464, 493-504
- Word length problems, in data processing equipment, 625-626
- WRU Searching Selector
- general description, 252-253
 - programming of searches, 256-260
- Wyandotte-ASTM punched card systems
- for spectral data, 183-207, 211-215, 216-223, 225-231
- X**erographic copying, and edge-punched or hand-sorted cards, 559, 561
- X-ray diffraction powder data, coded on punched cards, 207-215
- Z**atocoding system
- cards and equipment, 52-53, 341-345
 - random-like code patterns, 52-53, 344-346, 457 (footnote)
 - selector (sorting device), 52, 341-342
- Zatopleg code for chemical compounds, 490-491

RETURN **CHEMISTRY LIBRARY**
TO → 100 Hildebrand Hall 642-3753

LOAN PERIOD 1 7 DAYS	2	3
4	5	6

ALL BOOKS MAY BE RECALLED AFTER 7 DAYS
Renewable by telephone

DUE AS STAMPED BELOW

697



C034271525

Q131

A1222

1958

John L.

